

| Result No. | Query | | | DB | ID | Description |
|------------|--------|-------|--------|----|--------|----------------------|
| | Score | Match | Length | | | |
| 1 | 1232.5 | 37.5 | 623 | 3 | P79076 | P79076 coriolicus ve |
| 2 | 203 | 6.2 | 551 | 2 | Q9RH54 | Q9rh54 pantoea agg |
| 3 | 200.5 | 6.1 | 573 | 16 | Q9P190 | Q9p190 campylobact |
| 4 | 200 | 6.1 | 545 | 2 | Q9LBR8 | Q9lbr8 gluconobact |
| 5 | 191.5 | 5.8 | 615 | 2 | Q34214 | Q34214 pectobacter |
| 6 | 190.5 | 5.8 | 579 | 16 | Q9A716 | Q9a716 caulobacter |
| 7 | 182 | 5.5 | 591 | 16 | Q9I1K8 | Q9i1k8 pseudomonas |
| 8 | 180 | 5.5 | 553 | 2 | Q9XCR0 | Q9xcr0 pantoea cit |
| 9 | 179.5 | 5.5 | 529 | 17 | Q9HOR8 | Q9hqr8 halobacteri |
| 10 | 176 | 5.4 | 722 | 16 | Q9RZ26 | Q9rz26 deinococcus |
| 11 | 158.5 | 4.8 | 748 | 10 | Q9WOH4 | Q9wmh4 arabidopsis |
| 12 | 158.5 | 4.8 | 748 | 10 | Q9ABP3 | Q9abp3 arabidopsis |
| 13 | 155.5 | 4.7 | 499 | 16 | Q988P2 | Q988p2 rhizobium l |
| 14 | 155.5 | 4.7 | 769 | 3 | Q74253 | Q74253 pycnoporus |
| 15 | 149 | 4.5 | 479 | 16 | Q98D66 | Q98d66 rhizobium l |
| 16 | 141 | 4.3 | 523 | 16 | Q98C76 | Q98c76 rhizobium l |

Qy 591 HAIKSARSI 599

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2002, 11:07:23 ; Search time 13.49 Seconds
(without alignments)
1773.809 Million cell updates/sec

Title: US-09-856-327-2

Perfect score: 3284

Sequence: 1 MSLSTEQMLRDYPRSMQING.....IINTLKGDTGKNTGHRNL 618

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

* Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------------|----------------------|
| 1 | 126.5 | 3.9 | 773 | 1 CDH_PHACH | Q01738 phanerochaete |
| 2 | 116 | 3.5 | 514 | 1 GPD_MYCTU | O08407 mycobacteri |
| 3 | 107.5 | 3.3 | 964 | 1 YIN0_YEAST | P40467 saccharomyc |
| 4 | 103.5 | 3.2 | 612 | 1 AMYG_ASPOR | P36914 aspergillus |
| 5 | 103.5 | 3.2 | 639 | 1 AMYG_ASPAK | P23176 aspergillus |
| 6 | 102.5 | 3.1 | 1536 | 1 SIN3_YEAST | P22579 saccharomyc |
| 7 | 101 | 3.1 | 1010 | 1 SCA4_RICPA | O9a175 rickettsia |
| 8 | 100 | 3.0 | 663 | 1 ALOX_CANBO | Q00922 candida boi |
| 9 | 100 | 3.0 | 1142 | 1 ENAM_FIG | O97939 sus scrofa |
| 10 | 98.5 | 3.0 | 382 | 1 MTD_KLEPN | O9xbm6 klebsiella |
| 11 | 98.5 | 3.0 | 556 | 1 BETA_ECOLI | P17444 escherichia |
| 12 | 98 | 3.0 | 505 | 1 Y4NJ_RHISN | P55582 rhizobium s |
| 13 | 98 | 3.0 | 925 | 1 W70T_HUMAN | P57737 homo sapien |
| 14 | 98 | 3.0 | 1257 | 1 CCA4_BACTU | O45754 bacillus th |
| 15 | 96.5 | 2.9 | 684 | 1 Y492_MYCTU | Q11157 mycobacteri |
| 16 | 96 | 2.9 | 291 | 1 U502_HSV11 | P06485 herpes simp |
| 17 | 95.5 | 2.9 | 382 | 1 MTD_ECOLI | P09424 escherichia |
| 18 | 95.5 | 2.9 | 454 | 1 DLDH_RHOCA | P95596 rhodobacter |
| 19 | 95.5 | 2.9 | 519 | 1 AMVH_SACFI | P26989 saccharomyc |
| 20 | 95.5 | 2.9 | 708 | 1 HELS_BULSO | O97y99 sulfolobus |
| 21 | 95.5 | 2.9 | 883 | 1 RPOL_BPT7 | O05753 bacterioph |
| 22 | 95 | 2.9 | 487 | 1 ENGA_CHLNP | O9z762 chlamydia p |
| 23 | 95 | 2.9 | 500 | 1 YDAK_YEAST | P28817 saccharomyc |
| 24 | 94.5 | 2.9 | 855 | 1 GAF1_SCHPO | Q10280 schizosacch |
| 25 | 94.5 | 2.9 | 1211 | 1 BUN2_DROME | O24523 drosophila |
| 26 | 94.5 | 2.9 | 2193 | 1 POLG_CXAL6 | Q65900 c genome po |
| 27 | 94 | 2.9 | 606 | 1 PRIM_MYXIA | P50070 myxococcus |
| 28 | 94 | 2.9 | 886 | 1 SM6B_MOUSE | O54951 mus musculu |
| 29 | 93.5 | 2.8 | 406 | 1 YN05_YEAST | P53891 saccharomyc |
| 30 | 93.5 | 2.8 | 594 | 1 CIK1_YEAST | O01649 saccharomyc |
| 31 | 93.5 | 2.8 | 823 | 1 SCH9_YEAST | P11792 saccharomyc |
| 32 | 93.5 | 2.8 | 903 | 1 VGLB_HSV1F | P06436 herpes simp |
| 33 | 93.5 | 2.8 | 904 | 1 VGLB_HSV11 | P10211 herpes simp |

34 93.5 2.8 904 1 VGLB_HSV1P
35 93.5 2.8 1859 1 RP81_CAEEL
36 93 2.8 470 1 LEU2_AZOVI
37 93 2.8 552 1 CHOD_BREST
38 93 2.8 605 1 COX_ASPNG
39 93 2.8 737 1 AMY1_AEDAE
40 93 2.8 890 1 GLND_ECOLI
41 93 2.8 890 1 GLND_SALTY
42 93 2.8 953 1 YNM7_YEAST
43 93 2.8 955 1 VP2_BFV17
44 93 2.8 1012 1 UBAL_SCHPO
45 93 2.8 1117 1 CYT4_NEUCR

ALIGNMENTS

RESULT 1
CDH_PHACH STANDARD; PRT; 773 AA.
AC Q01738: 000047;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Cellobiose dehydrogenase precursor (EC 1.1.5.1) (CDH) (Cellobiose-quinone oxidoreductase).
GN CDH-1 AND CDH-2.
OS Phanerochaete chrysosporium.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Aphyllophorales; Corticiaceae; Phanerochaete.
OX NCBI_TaxID=5306;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OGC101;
RX MEDLINE=97077226; PubMed=8919793;
RA Li B., Nagalla S.R., Renganathan V.;
RT "Cloning of a cDNA encoding cellobiose dehydrogenase, a hemoflavoenzyme from Phanerochaete chrysosporium.";
RL Appl. Environ. Microbiol. 62:1329-1335(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=OGC101;
RX MEDLINE=97176414; PubMed=9023960;
RA Li B., Nagalla S.R., Renganathan V.;
RT "Cellobiose dehydrogenase from Phanerochaete chrysosporium is encoded by two allelic variants.";
RL Appl. Environ. Microbiol. 63:796-799(1997).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 19-208.
RX MEDLINE=20139694; PubMed=10673428;
RA Hallberg B.M., Bergfors T., Boeckbro K., Pettersson G., Henriksson G., Divine C.;
RT "A new scaffold for binding haem in the cytochrome domain of the extracellular flavocytochrome cellobiose dehydrogenase.";
RL Structure 8:79-88(2000).
CC -|- FUNCTION: DEGRADATES BOTH LIGNIN AND CELLULOSE. OXIDIZES CELLOBIOSE TO CELLOBIONOLACTONE.
CC -|- CATALYTIC ACTIVITY: Cellobiose + a quinone = cellobiono-1,5-lactone + a phenol.
CC -|- COFACTOR: ONE FAD AND ONE HEME B.
CC -|- SUBCELLULAR LOCATION: Secreted.
CC -|- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE GMC OXIDOREDUCTASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL; U46081; AAC49277.1; -

```
Query Match      3.5%; Score 116; DB 1; Length 514;
Best Local Similarity 22.8%; Pred. No. 0.34;
Matches 98; Conservative 48; Mismatches 163; Indels 120; Gaps 23;

QY 104 FNVVKGALQ--VSPVRNQNVTLPDQAWSPGSSAISNGKNPHOREFENLSAEAVTR 162
DB 77 FQGVYNAVHCRTFFRQONWDLAEG-FREVPQTFDDDDAFAQLAETLEKLDAAE--R 132

QY 163 GVGGMSTHTWCSTPRHPPM-ESL--PGIGREKLSNDAEDDKENNELYSEAEELIGTST 219
DB 133 GTGGHAYFLALPPKSPFVPCQLHKSLARPO-----GDRSVRVVIE----- 175

QY 220 KEFDESIRHTLVRLSLQADKDRQRIFFRPLPLACHRL--KNAPEYVEMHSAENLFHSYN 277
DB 176 KPFCHDLASARELNKAVNAVPEEAVER-----IDHVLCKETQNILARFANQLDFPIWN 231

QY 278 ----DDQKKLFTLTNHRCTRLATGGYEKKIGAAE--VRN-----LATRNPSSQL 324
DB 232 AHYVDHVO-----ITMAEDIGLGRAGYDYGIGAGVARDVIONHLMOLLALTAMEEPVSFH 285

QY 325 DSYIMA-KVYVLASGAIGNP-----QILYNSGFSG----- 353
DB 286 PAALQAEIKVLSTRLAEPDLDTTSRQYAGWGGGKVKVGLLDEEGFAEDSTTETFAA 345

QY 354 --LQVTPRNDLIP-----NLGRYITEQPMACQIVLRQEFV--DSVRDDPYGLPWWK 402
DB 346 ITLEVDTRRWAGVPEYLRGKRLGRVTEIALVF---RRAPHLPDFATMTDELG----- 396

QY 403 EAVAQHIKNTDALPIFRDEPOVTPFTTEHHPWHQIHRDA---FSYGAV-----GPE 455
DB 397 -----TNAMVIRVO-PDEGVTLRFGSKVPTAMEVRDVMNDFSYGSAFAEDSPE 444

QY 456 VDSRVIVDL 464
DB 445 AYERLILDV 453

RESULT 3
ID YINO YEAST STANDARD; PRT; 964 AA.
AC P40467;
* DT 01-FEB-1995 (Rel. 31, Created)
DT 16-FEB-1995 (Rel. 31, Last sequence update)
DE Putative 108.8 kDa transcriptional regulatory protein in FKH1-STH1
DE intergenic region.
DE Y11130W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Barrett B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: CONTAINS 1 ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR
CC CLUSTER DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC EMBL: Z38059; CAA86148.1;
DR PIR: S48404; S48404.
DR HSP: P08657; ICED.
DR SGB: S0001392; Y11130W.
DR InterPro: IPR001138; ZN2_Cy6_fungal.
DR Pfam: PF00172; ZN_Cy6; 1.
DR PRINTS: PR00054; FUNGALZNCYS.
DR SMART: SM00066; GAL4; 1.
DR PROSITE: PS00463; ZN2_Cy6_FUNGAL_1; 1.
DR PROSITE: PS00048; ZN2_Cy6_FUNGAL_2; 1.
KW Hypothetical protein; Transcription regulation; DNA-binding;
KW Nuclear protein; Zinc; Metal-binding.
FT DOMAIN 21 47 ZN(2)-CYS(6), FUNGAL-TYPE.
FT DOMAIN 811 896 HIS-RICH.
SQ SEQUENCE 964 AA; 108780 MW; AD5ABE59E4B022CC CRC64;

Query Match      3.3%; Score 107.5; DB 1; Length 964;
Best Local Similarity 19.5%; Pred. No. 3.8;
Matches 113; Conservative 86; Mismatches 187; Indels 193; Gaps 30;

QY 33 NDGVDFVTFAGSGPIGATYAK--LCVERGURVVMVEIGAADSFYVNAEGTAPYV-- 86
DB 322 DEGYKYFAARKLIDITNARDLNSIQAILMLII-----FLQCSARLSTCYTYIGVAM 373

QY 87 ----PGYHKK-----NEIEFQK-----DIDRFVNVTKGALQOVSPVRNQNVP 125
DB 374 RSLARAGFHRKLSNPGSPGPIEIMKRKLFYTIYKUDVINAMGLUPRSIPDDFQOTLP 433

QY 126 TLD-----PGAWSPGSAISNGKNPHOREF-----ENLSAEAVTRGVGGMST 169
DB 434 -LDLSDENITEVAYLPENQHSVLSSTGIS--NEHTKFLILNEIISELYPIKKTNIIS 489

QY 170 HWTCT--PRIHPMESLPGIGRPKLSN-DPAEDDKENNELYSEARLIGTSTKEFDESI 226
DB 490 HETVTSLELKLRLNWLDSLPLKELIPNAENIDPE-----YERANRLHL-----SF 533

QY 227 RHTLVRLSLQDAYKDRQRIREFL-----PLACHRLKN-----APE 261
DB 534 LHVQII-----LYRFFIHYLSRMNAENVPLCYRRARNSIAVARTVIKAKE 581

QY 262 YVE-----WHSAENLFHSIY-----NDKQKKLFTLTNHRCTRLALTGG 301
DB 582 MVSNNLLTGSYACYTIFYSVAGLLFYTHEAQLPKDSAREYDILKDAETGSRVLIQL 641

QY 302 YEKKGIAAEVNRLLATRNPSOLDSDYINAKVYVVLASGAINFQILYN--SGFSGLOVTPR 359
DB 642 KDSMAASRTYNLL-----NOIFEKLSNKTQLTA-----LHSSPSNESAFSLVTNN 687

QY 360 NDSLIPNLGRYITEQPMACQIVLRQEF-----VDSVRDDPYGLPWWKEAVAQHIKNPT 414
DB 688 SSALAPHLGDSL-QPPVFFSSQDTKNSFLAKSESTND-----YAMANYLNTPI 737

QY 415 DALPIPRDPEPV--TTPFTEHPHWTQIHRDAFSYGAVGPEVDSRVIVDLRWFEGATD 471
DB 738 SENPLNEAQQDOVQSGTGNMSNE-----RDPNNELSIIRLDNN----- 777

QY 472 PEANNLLVFQNDV---QDGYSMPQPT---FRYRPSTASN 504
DB 778 -GQSNILDATDDVFIRNDG-DIPTNSAFDFSSSKSNASN 814.

RESULT 4
ID AMYG_ASPOR STANDARD; PRT; 612 AA.
AC P36914;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Glucoamylase precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)
DE (1,4-alpha-D-glucan glucohydrolase).
GN GLAA.
```


| | |
|------------|--|
| RESULT | 11 |
| BETA_ECOLI | |
| ID | BETA_ECOLI STANDARD; PRT; 556 AA. |
| AC | PI7444; P77861; |
| DT | 01-AUG-1990 (Rel. 15, Created) |
| DT | 01-AUG-1990 (Rel. 15, Last sequence update) |
| DT | 16-OCT-2001 (Rel. 40, Last annotation update) |
| DE | Choline dehydrogenase (EC 1.1.1.99.1) (CHD). |
| GN | BETA OR B0311. |
| OS | Escherichia coli. |
| OC | Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; |
| OC | Escherichia |
| OX | NCBI_TaxID=562; |
| [1] | SEQUENCE FROM N.A. |
| RN | STRAIN-K12; |
| RC | MEDLINE-92065800; PubMed-1956285; |
| RX | Lamarck T., Kasen E., Eshoo M.W., Falkenberg P., McDougall J., |
| RA | Strom A.R.; |
| RA | "DNA sequence and analysis of the bet genes encoding the |
| RT | osmoregulatory choline-glycine betaine pathway of Escherichia coli."; |
| RT | Mol. Microbiol. 5:1049-1064(1991). |
| [2] | SEQUENCE FROM N.A. |
| RN | MEDLINE-91348527; PubMed-1879697; |
| RX | Boyd L.A., Adam L., Pelcher L.E., McHughen A., Hirji R., Selvaraj G.; |
| RA | "Characterization of an Escherichia coli gene encoding betaine |
| RA | aldehyde dehydrogenase (BADH): structural similarity to mammalian |
| RT | ALDHs and a plant BADH.; |
| RT | Gene 103:45-52(1991). |
| [3] | SEQUENCE FROM N.A. |
| RN | STRAIN-K12 / MG1655; |
| RC | MEDLINE-97426617; PubMed-9278503; |
| RX | Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., |
| RA | Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., |
| RA | Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., |
| RA | Mau B., Shao Y.; |
| RA | "The complete genome sequence of Escherichia coli K-12."; |
| RT | Science 277:1453-1474(1997). |
| [4] | SEQUENCE FROM N.A. |
| RN | Duncan M., Allen E., Araujo R., Aparicio A.M., Chung E., Davis K., |
| RA | Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D., |
| RA | Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.; |
| RL | Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases |

. Db 108 PGQOALPSAGVVLG--PEDLPVEVLQFHTSDGILVSAGTVKVVWDAKQO---PLT 162
QY 184 SLPGIGRPKLSNPAEDDKENNELYSSEAERLIGTSTKEFDESIRHTVLRLSLQDAYKDRQ 243
Db 163 EL-----AAHGLVQSAVWSRQALVCT-----ACKDKQ 191
QY 244 -RIFRPLPLACHRLK-NAPEYVEWH-----SAENLPHSYNDKQKLL----- 284
Db 192 LRIFDP-----RTKPRASOSTOAHNSRSLRAWMTWHLVSTGTGFNQMREREVLWD 245
QY 285 -----FTLLTNIRCT-----RLALGGYKKEKIGAAEVRNLLATRNPSQ--L 324
Db 246 RFESSALASITLSTGLCLPLDPSGLLVLAGKERQLYCEVYVQPALSPVTCVL 305
QY 325 DSIMAKVYVLAGATGNPOILYNSFGSLQVTPRNDSLIPNLGRYITEOPMAFCQIVLR 384
Db 306 ES-----VLRGAALVPRQALAVMGCEVLRVLQSLDAITVPGYHVRKAVEFHEDL-- 356
QY 385 QEFVDSV---RDDPYGLPW--KEAVAQHIKNP-----TDALPIPFDRPEQVIT 430
Db 357 --FPDTAGCVATDPHG--WAGDNQOVQVSLNPACRPHPSFSL-VPPAEPLPDTAQ 411
QY 431 PFTEHPWHTQIHRDAFSYGAVGPEVDSRVIVDLRFWFGATDPEANNLLVFQNDVQDGYM 490
Db 412 PAYMETP-----PSSLTSPSPSSL 444
QY 491 PQTFYRSTASNV 505
Db 431 P-PSSLTSPSPSSL 444

RESULT 14
CCAA_BACTU STANDARD; PRT; 1257 AA.
AC Q45754;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pesticidal crystal protein cry12Aa (insecticidal delta-endotoxin
DE CryIIA(a)) (crystalline entomocidal protoxin) (142 kDa crystal
DE protein).
GN CRY12AA OR CRYIIA(A) OR CRYVB.
OS Bacillus thuringiensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1428;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 2-11.
RC STRAIN=NRRL B-18244 / PS33F2;
RA Narva K.E., Payne J.M., Schwab G.E., Hickie L.A., Galasan T.,
RA Sick A.J.;
RT "Novel Bacillus thuringiensis microbes active against nematodes, and
RT genes encoding novel nematode-active toxins cloned from Bacillus
RT thuringi.";
RL Patent number EP0462721, 27-DEC-1991.
CC -1- FUNCTION: ENDOTOXIN WITH NEMATOCIDAL ACTIVITY.
CC -1- DEVELOPMENTAL STAGE: THE CRYSTAL PROTEIN IS PRODUCED DURING
CC SPOULATION AND IS ACCUMULATED BOTH AS AN INCLUSION AND AS PART
CC OF THE SPORE COAT.
CC -1- MISCELLANEOUS: TOXIC SEGMENT OF THE PROTEIN IS LOCATED IN THE
CC N-TERMINUS.
CC -1- SIMILARITY: BELONGS TO THE DELTA ENDOTOXIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L07027; AAA22355.1;
DR InterPro: IPR001178; Endotoxin.

DR Pfam: PF00555; endotoxin; 2.
KW Toxin; Sporulation.
SQ SEQUENCE 1257 AA; 142265 MW; 3D9888FFC6C0E3981 CRC64;

Query Match 3.0%; Score 98; DB 1; Length 1257;
Best Local Similarity 20.3%; Pred. No. 30;
Matches 132; Conservative 88; Mismatches 243; Indels 188; Gaps 36;

QY 15 SMOINGOIPKNAIHEHYGVNDGVDFIAGSPIGATYAKLCVEAGLRVVMVEIGAADSFYA 74
Db 300 SSPIYQVPKMNQNTSSIVPSDLFHYQGLVKLEFSTRDNDGLAKIFT--GRNTFYK 357
QY 75 VNAECTAVPYVPGYHKKNEIEFKDIDREVNVIKQALQVQVVRNQNYPT-LDPAWS 133
Db 358 -----SPNTHYHVDFSYNTQSSGNISRGSSNIPIDLNPIIISTCIRNSFYK 406
QY 134 APGSSAISNGKPNHQREFENLSAEAVTRGVGGMSTHTWCSTPRIHPMESLPGIGRPL 193
Db 407 AIAGSSVLVNFKDGTO---GYAFAQAAPTGA-----WD-----HSFIES----- 442
QY 194 SNDPAEDDKENNELYSSEAERLIGTSTKEFDESIRHTLV-LRSLQDAYKDRIFRPLPLA 252
Db 443 --DGAPEGHKLNIYITSP---GDTLDFINV--YLISTPTINELSTEIKGF---PAE 491
QY 253 CHRLKN-----APEYVEWHSAENLFHSIYNDKOKKFLTLTNHRCRTRALTGGEYK 304
Db 492 KGYIKQGIKMYGKPEYINGAQPVNL-----ENQOTLIFE-----HASKTAQYTI 538
QY 305 KIGAAEVRNL-----LATNPSSQLDSYIMAKVYVLAGAIGNPOILYNSFGSL 354
Db 539 RIRYASTQGTGKYFRLDNQLPTSHNGYV-----TGNIGENDLYTIG--SY 588
QY 355 QVTPRNDSLIPNLGRYTEOPMAFCQI-----VL-ROEFV--DSVRDDPYGLPWKE 403
Db 589 TITEGHTL-----QIOHNDKNGWLDRIEFVKDSLQDSP----- 624
QY 404 AVAQHAKNPTDALPIFRDPEQV---TTPFTEHP--W-----HTQIHRDAFSYGAV 452
Db 625 -----QDSPEVHESTIIFDKSSPTIWSNNKHSYSHILEG--SYTSQ 665
QY 453 GPEVDSRVIVDLRFWFGATDPEANNLLVFQNDVQDGYMPOPTFRYRSTASNVARKMM 511
Db 666 G-SYPHNLINL--FHPTDPRNHTTHVNGDMNVDTG-----KDSVADGLNFKIT 714
QY 512 ADMCEVASNLG-----GYLPTSPPOFMDPGLAHLAG-TTRIGFKAT-----TVA 556
Db 715 AITPSDAWYSGTITSMHLENDNNFKITPKF---ELSNELNITTVQNALFASQAQDTLA 771
QY 557 DNNSLVMDPANLYVAGNGTIRTFG-ENPTLTSMCHAIKSARSIINTLKG 606
Db 772 SNVSDYW-IEQVVMKVDALSDEVFGEKRALKLVNQAKRLSKIRNLLIG 821

RESULT 15
I492_MYCTU
ID Y492_MYCTU STANDARD; PRT; 684 AA.
AC Q11157;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 57.3 kDa protein GMC-type oxidoreductase RV0492c.
GN RV0492C OR MT0511/MT0512 OR MTCY2069.18C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2002, 11:05:57 ; Search time 20.92 Seconds
(without alignments)
2838.585 Million cell updates/sec

Title: US-09-856-327-2
Perfect score: 3284
Sequence: 1 MSLSTEQMLRDYPRSMQING.....IINTLKGGTGDKNKGHRNL 618

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR-71.*

1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------|--------------------|
| 1 | 200.5 | 6.1 | 573 | E81385 | probable oxidoredu |
| 2 | 191.5 | 5.8 | 615 | B38575 | gluconate 2-dehydr |
| 3 | 190.5 | 5.8 | 579 | H87451 | oxidoreductase, GM |
| 4 | 187.5 | 5.7 | 619 | A82003 | dehydrogenase chai |
| 5 | 182 | 5.5 | 591 | H83362 | gluconate dehydrog |
| 6 | 179.5 | 5.5 | 529 | A84260 | hypothetical prote |
| 7 | 176 | 5.4 | 722 | B75608 | GMC oxidoreductase |
| 8 | 171.5 | 5.2 | 562 | JC7628 | glucoside 3-dehydr |
| 9 | 158.5 | 4.8 | 748 | T10651 | hypothetical prote |
| 10 | 148.5 | 4.5 | 561 | E98192 | probable oxidoredu |
| 11 | 148.5 | 4.5 | 561 | A83094 | oxidoreductase Atu |
| 12 | 138.5 | 4.2 | 527 | E98202 | dehydrogenase chai |
| 13 | 138.5 | 4.2 | 527 | AD3084 | dehydrogenase Atu4 |
| 14 | 138.5 | 4.2 | 768 | JC6584 | cellobiose oxidase |
| 15 | 138 | 4.2 | 494 | AD3255 | l-sorbose dehydrog |
| 16 | 130.5 | 4.0 | 770 | S60676 | cellobiose oxidase |
| 17 | 127.5 | 3.9 | 578 | F70736 | probable chob prot |
| 18 | 124.5 | 3.8 | 578 | D95400 | probable oxidoredu |
| 19 | 123 | 3.7 | 531 | B83600 | probable oxidoredu |
| 20 | 121 | 3.7 | 1498 | AF1082 | B. subtilis Yuka p |
| 21 | 117 | 3.6 | 736 | A86171 | hypothetical prote |
| 22 | 116 | 3.5 | 502 | T18562 | hypothetical prote |
| 23 | 116 | 3.5 | 514 | B70917 | probable zwf2 prot |
| 24 | 116 | 3.5 | 599 | T19711 | hypothetical prote |
| 25 | 114 | 3.5 | 1012 | T00958 | hypothetical prote |
| 26 | 113.5 | 3.5 | 678 | T05821 | hypothetical prote |
| 27 | 111.5 | 3.4 | 1042 | T16169 | hypothetical prote |
| 28 | 110.5 | 3.4 | 589 | T50698 | probable mandeloni |
| 29 | 110.5 | 3.4 | 5170 | T15348 | hypothetical prote |

| | | | | | | |
|----|-------|-----|------|---|--------|--------------------|
| 30 | 109.5 | 3.3 | 1125 | 2 | T19193 | hypothetical prote |
| 31 | 108 | 3.3 | 1498 | 2 | AG1439 | B. subtilis Yuka p |
| 32 | 107.5 | 3.3 | 557 | 2 | F83181 | probable GMC-type |
| 33 | 107.5 | 3.3 | 567 | 2 | AC0143 | choline dehydrogen |
| 34 | 107.5 | 3.3 | 964 | 2 | S48404 | probable membrane |
| 35 | 107 | 3.3 | 1018 | 2 | T40253 | hypothetical prote |
| 36 | 106.5 | 3.2 | 775 | 2 | B72074 | hypothetical prote |
| 37 | 106.5 | 3.2 | 775 | 2 | C81594 | hypothetical prote |
| 38 | 106.5 | 3.2 | 775 | 2 | D86549 | hypothetical prote |
| 39 | 106.5 | 3.2 | 832 | 2 | H84848 | phospholipase D [i |
| 40 | 105.5 | 3.2 | 1611 | 2 | T38236 | hypothetical prote |
| 41 | 105 | 3.2 | 499 | 2 | T45749 | hypothetical prote |
| 42 | 104 | 3.2 | 509 | 2 | D87452 | oxidoreductase, GM |
| 43 | 104 | 3.2 | 585 | 2 | S72824 | cholesterd1 oxidas |
| 44 | 104 | 3.2 | 986 | 2 | T33135 | hypothetical prote |
| 45 | 104 | 3.2 | 1917 | 2 | C88728 | protein C48A7.1 [1 |

ALIGNMENTS

RESULT 1

E81385
probable oxidoreductase chain Cj0415 [imported] - Campylobacter jejuni (strain NCTC 11637)
C:Species: Campylobacter jejuni.
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000
C:Accession: E81385
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chli
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; B
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
A:Reference number: A81250; MUID:20150912
A:Accession: E81385
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-573 <PAB>
A:Cross-references: GB:AL139075; GB:AL111168; NID:g6967817; PIDN:CAB74251.1; PID:g696
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: Cj0415

Query Match 6.1%; Score 200.5; DB 2; Length 573;
Best Local Similarity 20.6%; Pred. No. 3.5e-07;
Matches 135; Conservative 74; Mismatches 270; Indels 175; Gaps 26;

QY 36 VDVFTAGSPICATYAKLCVEAGLRVWVEIG---AADSFYAVNAEGTAVPVGYHKK 92

Db 8 VDVVTGAGWTGGIIVAAELTKAGLVLSLEIGHMQSTENFNHDEWRYGINY----- 60

QY 93 NEIEFQKOIDRFVNVIKGALQOVS---VPVRNQNVPVTLDPGAWSPAPGSSAISNGKNPHQ 149

Db 61 -----GLMQDCSKDVTFRH-----DPSGLALP----- 83

QY 150 REFENLSAEAVTRGVGGMSTHTWCSTPRTHPPMESLFGICRP-----KLNDPAEDD----- 201

Db 84 --YRKMGSLFLNNYGGAGVHWNGWTFRPMYPDFEIQTLKQRYGNKLGNDYTLQDWGVT 141

QY 202 -KENNELYSEARLIGTSTKEF-----DESIRHTVLVRLSLQDAYKQRORI 245

Db 142 YKDMPEYVDRFEKTCGSGVEPNPLAEKMGAFSSPYQPPELNTKMLKRFESAASNNLH 201

QY 246 FRPLPLA-----CHRLKNAPEYVEHSAENLFHSIYNDKQKK 283

Db 202 TYRLPASNSKGYTPNDGODLAPCOYCAICREFG-----CEYGAKASPLNTVTPKAMSTG 256

QY 284 LFTLTNRHCRTRALATGGYEKKIGAAEVENLLATRNPSQLDSYIM-AKVYVLASGAIGN 342

Db 257 KVTIRYSNVTQIL-----KKDGKVTGVKFDVDTRI-----MKEYIQAPADIVLVTSMFNN 306

QY 343 POILYNSFGSLQVTP-----RNDSLIPNLG--RYITEQPMAPCOVLVLRQEFVDSVR 392

Db 307 AKLLMVSNI-GEQYDPKTKGTLGRNYCYQMNGMTTAFDFEQNTF-----MGSGALGTTTS 361

Db 343 DKLGITFVKA-----ALNLSPEH-----YADAKAGHAGFGDQTGAKQGTILNTFLKDAFEAG 393
QY 287 --LLTNHRCRLALTGGEYKKGAEVRLNLLATRNSSQLDSYIMAKVYVVLASGAINPQ 344
Db 394 ARILVTRAQIRVLVDG-----RAAGVSATVTMGDETROI--VRAPOVYVAGGALETPA 446
QY 345 ILNYSFGSLQVTPRNDLSIPNLGRYITEOPMAFCOIVLRQEFVDSVRDDPYG---LPWW 401
Db 447 LLLRSIGG-----PAAGRYLRHPAGLVAGI-----YGEDQRAWW 482
QY 402 KEVAOHIAKNPTDALPIPRDE-----PQVTT--PFT--EEH-PWHTQI 442
Db 483 GP-----POSGILKOFADHENGHGTIEGVOYGPALMASGLPWTGGEARHDLMSKF 533
QY 443 HRDAFSYGAVPEVDSRVIVD-----LRWFGATDP--EANNLLVFONDVODGYSM-----P 491
Db 534 HRWATFVSIVODRGHGVTVDDGNAVHTYALTDLDLARN---FRGVTESIRLHEAAGA 590
QY 492 QPTFRYRSTASNVARKMADCEVAS---NLGGYLPTSPPOFMDPGLALHLAGTRIG 548
Db 591 EEIVALAPGVAPWRGDDLEAFIGVQAQVPLGAGGTVFS-----AHQMGSRWG 640
QY 549 FDKATVADNNSLVDFANLYVAGNTIRTFGENPTLISMCHAISRIINTLKGTD 608
Db 641 SDPQTSVADPDGQLHDVPGVWIGDTSFAFTCSGVNPNVSCMALASRTAEKLLAAMEGADG 700
QY 609 GKNTG 613
Db 701 GTGSG 705

RESULT 8

JC7628
glucoside 3-dehydrogenase (EC 1.1.99.13) - Halomonas sp. alpha-15
C:Species: Halomonas sp. alpha-15
C>Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 30-Jun-2001
C:Accession: JC7628; PC7123
R:Kojima, K.; Tsugawa, W.; Sode, K.
Biochem. Biophys. Res. Commun. 282, 21-27, 2001
A:Title: Cloning and expression of glucose 3-dehydrogenase from Halomonas sp. alpha-15
A:Reference number: JC7628; MUID:21164693; PMID:11263965
A:Accession: JC7628
A:Molecule type: DNA
A:Residues: 1-562 <KOJ>
A:Accession: PC7123
A:Molecule type: protein
A:Residues: 2-11 <KO2>
C:Comment: This enzyme, as an oligomeric enzyme composed of catalytic and electron trans
and disaccharides.
C:Genetics:
A:Gene: g3dh
C:Keywords: oxidoreductase

Query Match 5.2%; Score 171.5; DB 2; Length 562;
Best Local Similarity 19.9%; Pred. No. 5.7e-05;
Matches 133; Conservative 85; Mismatches 236; Indels 215; Gaps 32;

QY 37 DVFIAGSGPIGATYAKLCVEAGLRVYVMEIGAADSFYAVNAEEGTAVPYVGYHK----- 91
Db 7 DAIVGSGISGGWAAKELTEKGLVLL-----ERGRNIEHVKYDHNADKEA 53
QY 92 -----KNETEFKIDREFNVNKGALQVSVVVRNQNVPITLDP--CAWSAPPGSSAISNG 144
Db 54 WDPHPRNE-PTQMIKY-----PVLKRDYPLNEATLGMWA-----DEQ 91
QY 145 KNPH--QREFENLSAEAVTRG--VGMSTHTWCSTPRIHPHPSLPGIGRPKLSND---P 197
Db 92 ANPYVEKRFDPW-----RGYHVGRSLWGRQSYRLSP--MDFEANQREGAIDWP 143
QY 198 AEDDKENNELYSAEERLIG--TSTKEFDESIRHTLVLSRLQDAYKQRIFRPLPLACHRL 256

Db 144 YEDLAPW---YDYVERFAGIAGTQE-----GLDILPDG-----EFLPPIPLNC--- 183
QY 257 KNAPEVVEHSAENLPHSIYNDKQKLFLLTNHRCRLALTGALGGEYKKGAEVRLNLLA 316
Db 184 -----VEDDAKRI-----KEAF-----GGORHLHSRVANITQ 212
QY 317 TRNPSQLDSYIMAKVYVVLASGAIGNPQILYNSGSLQVTPRNDLSIPNLGRYITEOPM 376
Db 213 PRPEQNRVNCQYRNKCW-----LGCP---YGAVFSTQSATLPAAVATGNL---TLRPF 259
QY 377 AFCQIVL---RO-----EFVDS-----VRDDPYGLPW-----WKEA 404
Db 260 SIYSQVLYDKDRQARGVEVIDAETHVEHYETADVIFLNASTFTNTWILNASTADVWEGG 319
QY 405 VAQHIKAKNPTDALPIPR---DPEQVITPTTEHPWHTQIHR-----DAF 447
Db 320 LGSSSELGHNVMDHFRGASGEVGYLDKYYFGRPAGFYIPRFRNVGDEQBSYVRGP 379
QY 448 SY-GAYGPEYDSRVIVDLR-----W-----FGATDPEANNLLVFQNDVOD 486
Db 380 GYGASREGWDREIAELNIGADLKQALTOPGGWTIGTGTGEMLPDHDNRISLDHSVRD 439
QY 487 GYSMPQPTFRYRSTASNVARKMADMBCE-----VASNLGGYLPTSPPOFMDPGL 537
Db 440 KMGLPVLSIDVELKQNERDMRDVQDADVLEAAGVKNVKGVDGYA-----PGM 490
QY 538 ALHLACTTRIGDPKATTVADNNSLVDFANLYVAGNTIRTFGENPTLISMCHAISAR 597
Db 491 GIHEMGTARMGRDPKTSVLNSHQVNDAPNVFVDGACMTSSSCVNPSTLTMALTARAYD 550
QY 598 SIINTLKG 606
Db 551 YAVEELKRG 559

RESULT 9

T10651
hypothetical protein T5F17.20 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Sep-2000
C:Accession: T10651
R:Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Ba
submitted to the Protein Sequence Database, June 1999
A:Reference number: Z16533
A:Accession: T10651
A:Molecule type: DNA
A:Residues: 1-748 <BEV>
A:Cross-references: EMBL:AL049917; GSPDB:GN00062; ATSP:T5F17.20
A:Experimental source: cultivar Columbia; BAC clone T5F17
C:Genetics:
A:Gene: ATSP:T5F17.20
A:Map position: 4
A:Introns: 70/3
C:Superfamily: Arabidopsis thaliana hypothetical protein T5K18.160

Query Match 4.8%; Score 158.5; DB 2; Length 748;
Best Local Similarity 19.4%; Pred. No. 0.00089;
Matches 126; Conservative 90; Mismatches 215; Indels 217; Gaps 28;

QY 37 DVFIAGSGPIGATYAKLCVEAGLRVYVMEIGAADSFYAVNAEEGTAVPYVGYHKNEIE 96
Db 239 DAVVVGSGGGVAAANLAKAGLVLEKG---NYTADHYSGLVPSMLELYEKG--- 292
QY 97 FKQIDRFNVNKGALQVSVVVRNQNVPITLDPGANSAPPGSSAISNGKPHOREFNLS 156
Db 293 -----GLLTV-----DGK-----FMLLA 306
QY 157 AEAVTRGVGMSHTWCSTPRIHPHPSLPGIGRPKLSNDPAEDDKENNELYSAEERLIG 216
Db 307 GSAVG---GGTAVNWSAS---INTPDHVL-----QEW-----SEGSKIKF 340
QY 217 TSTKEF---DE-SIRHTLVLSRLQDAYKQRIFRPLPLACHRL-----KNAPE-- 261

| | | | |
|---|-----|---|-----|
| Db | 109 | VEH-----EDGVSPAMPVSAEAPYAEARLEFV-ROGAGDDTPEPPRSPAYMH | 158 |
| Qy | 209 | SEA---ERLIGTSTKEFDESIRHTLVLSLQSDAYKDRORIFRPLPLA-----CHRLKN | 258 |
| Db | 159 | APIPHEPVIGRVAKGFER-----LGLRPFH-----MPSAIDYGGPGLCHRCGT | 201 |
| Qy | 259 | APERYVHSAENLPHSIYNDOKKKLF-----TLTNNHRCRTIALTGGYEKKIGAA | 309 |
| Db | 202 | CDAFVCRFDAG-----DAETRLLRPALRHPNVSLLTGARVRRL-TADGDKHIVAV | 252 |
| Qy | 310 | EVNRLLATRNPSOLDGVINAKVYVLASGAIGNPOILYNSGFSGLQVTPRNDLSLPN--- | 366 |
| Db | 253 | EIER-----AGEITT-IEAPLEFVLSAGAINLSALILRSA-----DEKKENGLA | 294 |
| Qy | 367 | LCRYITEQPM-----FCQIVLRQEEFVSVRDDP----- | 395 |
| Db | 295 | NSSGVGWYRILMNHLSGLMGLLPFTINDTRFPKTMSLNDFDCTPGDEAARGNVQMLGNI | 354 |
| Qy | 396 | -----VGLPWKEAVAQHAOKNPTDALPIFPDRDPEQVTPPTTEHPHWHQIHRDAFSY | 449 |
| Db | 355 | QGPMIRAAYPMMRPLLANLLARHSVDPLVM-----SED----- | 387 |
| Qy | 450 | GAVGPEVDSRVIVDLRFEGATDPEANLLVFQNDVQDYSMPQPTFRYRPSTASN--- | 505 |
| Db | 388 | -----TPKYDSRV-----KFWKNGAE-----LIYRPGDREAHOREV | 419 |
| Qy | 506 | RARKWADMCVEASNLGGYLTSPPOFMDPGLALHLAGTTRIGFDKATTVAONNSLVWDF | 565 |
| Db | 420 | RHMSLLRKKNGFPVVLGHSFGIEAPS-----HCGGIVRMGDDPKAALNALCQTYDH | 471 |
| Qy | 566 | ANLYVAGNGTIRTGFGNGENPLTSMCHAUK | 594 |
| Db | 472 | PNLYVVDAGFFPSSAALNPLALTVAQAALR | 500 |
| RESULT 13 | | | |
| AD3084 | | | |
| dehydrogenase Atu4296 [imported] - Agrobacterium tumefaciens (strain C58, I | | | |
| C:Species: Agrobacterium tumefaciens | | | |
| C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002 | | | |
| C:Accession: AD3084 | | | |
| R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Ch | | | |
| erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; L | | | |
| ; Karp, P.; Romero, P.; Zhang, S. | | | |
| Science 294, 2317-2323, 2001 | | | |
| A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; | | | |
| ster, E.W. | | | |
| A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaci | | | |
| A:Reference number: AB2577; PMID:11743193 | | | |
| A:Accession: AD3084 | | | |
| A>Status: preliminary | | | |
| A:Molecule type: DNA | | | |
| A:Residues: 1-527 <KUR> | | | |
| A:Cross-references: GB:AE008689; PIDN:RAL45090.1; PID:gl1742757; GSPDB:GN00 | | | |
| A:Experimental source: strain C58 (Dupont) | | | |
| C:Genetics: | | | |
| A:Gene: Atu4296 | | | |
| A:Map position: linear chromosome | | | |

| | | | | |
|-----------------------|------------------|------------------|-------------|-------------|
| Query Match | 4.28; | Score 138.5; | DB 2; | Length 527; |
| Best Local Similarity | 19.18; | Pred. No. 0.017; | | |
| Matches 120; | Conservative 74; | Mismatches 222; | Indels 213; | Gaps 25; |

| | Matches | 120; | Conservative | 74; | Mismatches | 222; | Indels | 213; | Gaps |
|----|---------|------------------------------------|-------------------------------|-----|------------|------|--------|------|------|
| Qy | 37 | DVFIAGSGPIGATYAKLCVEAGLRVVMVIGAA | DSFYAVNAEGTAVPVPGYHKKHKEIE | 96 | | | | | |
| | | ! : ! ! ! ! ! : | ! : ! : ! : ! : | | | | | | |
| Db | 14 | DIVIGSGVGSVALKLAATGAKILLIERGEKLPNE | PENAD--AAEAVFVQNRYETD | 70 | | | | | |
| | | ! : ! : ! : ! : | ! : ! : ! : ! : | | | | | | |
| Qy | 97 | FOKIDREVMNVIKALQQVSVPRNQNVPTLDPG | AWSAAPP-----SSAISNGKNPHORE | 151 | | | | | |
| | | ! : ! : ! : ! : | ! : ! : ! : ! : | | | | | | |
| Db | 71 | -----YRDETGTATPGQYVYVGGHKTFCYGT | AMFRFRDRDRPRE | 108 | | | | | |
| | | ! : ! : ! : ! : | ! : ! : ! : ! : | | | | | | |
| Qy | 152 | FENLSAEAVTRGVGGMSTHTWCSTPRIHP--- | PMESLFCIGRKLNSNDPAEDDKKENNELY | 208 | | | | | |

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2002, 11:03:42 ; Search time 14.68 Seconds
(without alignments)
1028.271 Million cell updates/sec

Title: US-09-856-327-2
Perfect score: 3284
Sequence: 1 MSLSTEQLRDPYPRSMQING.....IINTLKGTGDKNTGEHRNL 618

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA:*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query | Score | Match | Length | DB ID | Description |
|------------|--------|-------|-------|--------|-------------------|--------------------|
| 1 | 1232.5 | 37.5 | 623 | 1 | US-08-734-925-2 | Sequence 2, Appli |
| 2 | 1210 | 36.8 | 622 | 4 | US-09-305-381-2 | Sequence 2, Appli |
| 3 | 1091 | 33.2 | 566 | 4 | US-09-023-731-1 | Sequence 1, Appli |
| 4 | 192 | 5.8 | 51 | 4 | US-09-023-731-12 | Sequence 12, Appli |
| 5 | 191.5 | 5.8 | 615 | 4 | US-09-297-937C-9 | Sequence 9, Appli |
| 6 | 191.5 | 5.8 | 1276 | 4 | US-09-297-937C-13 | Sequence 13, Appli |
| 7 | 130.5 | 4.0 | 58 | 4 | US-09-023-731-11 | Sequence 11, Appli |
| 8 | 118.5 | 3.6 | 785 | 3 | US-09-265-108-2 | Sequence 2, Appli |
| 9 | 118.5 | 3.6 | 785 | 4 | US-09-479-264-2 | Sequence 2, Appli |
| 10 | 115 | 3.5 | 50 | 4 | US-09-023-731-14 | Sequence 14, Appli |
| 11 | 108 | 3.3 | 35 | 4 | US-09-023-731-15 | Sequence 15, Appli |
| 12 | 104.5 | 3.2 | 1876 | 2 | US-08-609-049A-12 | Sequence 12, Appli |
| 13 | 104.5 | 3.2 | 1876 | 4 | US-09-170-996-12 | Sequence 12, Appli |
| 14 | 102.5 | 3.1 | 61 | 4 | US-09-023-731-6 | Sequence 6, Appli |
| 15 | 100.5 | 3.1 | 806 | 3 | US-08-549-515-5 | Sequence 5, Appli |
| 16 | 100.5 | 3.1 | 806 | 3 | US-08-549-515-11 | Sequence 11, Appli |
| 17 | 100 | 3.0 | 664 | 1 | US-08-485-284A-5 | Sequence 5, Appli |
| 18 | 98.5 | 3.0 | 543 | 3 | US-09-199-229-2 | Sequence 2, Appli |
| 19 | 98.5 | 3.0 | 543 | 4 | US-09-443-087-2 | Sequence 2, Appli |
| 20 | 98.5 | 3.0 | 543 | 4 | US-09-687-298-2 | Sequence 2, Appli |
| 21 | 98 | 3.0 | 1257 | 1 | US-08-049-783-2 | Sequence 2, Appli |
| 22 | 98 | 3.0 | 1257 | 1 | US-08-158-232-6 | Sequence 6, Appli |
| 23 | 98 | 3.0 | 1257 | 1 | US-08-304-628-6 | Sequence 6, Appli |
| 24 | 98 | 3.0 | 1257 | 1 | US-08-316-301A-6 | Sequence 6, Appli |
| 25 | 98 | 3.0 | 1257 | 2 | US-08-611-928-6 | Sequence 6, Appli |
| 26 | 98 | 3.0 | 1257 | 3 | US-09-173-891-6 | Sequence 6, Appli |
| 27 | 98 | 3.0 | 1257 | 4 | US-09-076-137-6 | Sequence 6, Appli |

Sequence 6, Appli
Sequence 5, Appli
Sequence 5, Appli
Sequence 28, Appli
Sequence 28, Appli
Sequence 2, Appli
Sequence 22, Appli
Sequence 22, Appli
Sequence 1, Appli
Sequence 1, Appli
Patent No. 5244792
Sequence 32, Appli
Sequence 30, Appli
Sequence 2, Appli
Sequence 14, Appli
Sequence 2, Appli
Sequence 6, Appli
Sequence 6, Appli

28 98 3.0 1257 5 PCT-US92-03624-6
29 97 3.0 3170 2 US-07-642-734C-5
30 97 3.0 3170 3 US-08-439-009A-5
31 96.5 2.9 1876 2 US-08-609-049A-28
32 96.5 2.9 1876 4 US-09-170-996-28
33 95.5 2.9 382 1 US-08-186-833-2
34 93.5 2.8 903 3 US-08-804-439A-22
35 93.5 2.8 903 3 US-08-720-229-22
36 93.5 2.8 904 4 US-08-632-537-1
37 93.5 2.8 904 5 PCT-US96-05316-1
38 93.5 2.8 904 6 5244792-4
39 93 2.8 604 2 US-08-746-283-32
40 93 2.8 604 2 US-08-746-257A-30
41 93 2.8 605 1 US-08-333-802-2
42 93 2.8 4545 2 US-08-804-227C-14
43 92.5 2.8 4472 2 US-08-804-227C-2
44 92 2.8 844 1 US-07-731-157A-6
45 92 2.8 844 2 US-08-541-780-6

ALIGNMENTS

RESULT 1
US-08-734-925-2
; Sequence 2, Application US/08734925
; Patent No. 5712139
; GENERAL INFORMATION:
; APPLICANT: NISHIMURA, IKUKO
; APPLICANT: OKADA, KIMIHARU
; APPLICANT: MINAMIHARA, TOMOYUKI
; APPLICANT: KAWAI, GENSUHIRO
; APPLICANT: KAYAMA, YASUJI
; APPLICANT: SUZUKI, MASARU
; TITLE OF INVENTION: NOVEL PYRANOSE OXIDASE, PYRANOSE OXIDASE
; TITLE OF INVENTION: GENE, NOVEL RECOMBINANT DNA AND PROCESS FOR PRODUCING
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/734,925
FILING DATE: 22-OCT-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/568,428
FILING DATE: 06-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 7126-002-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 623 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-734-925-2

Patent No. 6291648

GENERAL INFORMATION:

APPLICANT: Kawamura, Yukio; Morita,
APPLICANT: Akhihiro; Izumo, Koji.; Saka, Tomohide.
TITLE OF INVENTION: ANTITUMOR PROTEIN AND
TITLE OF INVENTION: GENE ENCODING SAME
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154

COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS.
SOFTWARE: MICROSOFT WORD 97

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,731
FILING DATE: 13-FEB-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 29275/1997
FILING DATE: 13-FEB-1997

ATTORNEY/AGENT INFORMATION:
NAME: KENNETH H. SONNENFELD
REGISTRATION NUMBER: 33,285
REFERENCE/DOCKET NUMBER: 3479-4000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792

INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 51
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide

US-09-023-731-12

```

; ; FILING DATE: 13-FEB-1997
; ; ATTORNEY/AGENT INFORMATION:
; ; NAME: KENNETH H. SONNENFELD
; ; REGISTRATION NUMBER: 33,285
; ; REFERENCE/DOCKET NUMBER: 3479-4000
; ; TELECOMMUNICATION INFORMATION:
; ; TELEPHONE: (212) 758-4800
; ; TELEFAX: (212) 751-6849
; ; TELEX: 421792
; ; INFORMATION FOR SEQ ID NO: 12:
; ; SEQUENCE CHARACTERISTICS:
; ; LENGTH: 51
; ; TYPE: amino acid
; ; TOPOLOGY: linear
; ; MOLECULE TYPE: peptide
; ;
; ; US-09-023-731-12
;
; Query Match 5.8%; Score 192; DB 4; Length 51;
; Best Local Similarity 62.7%; Pred. No. 1e-12;
; Matches 32; Conservative 6; Mismatches 13; Indels 0; Gaps 0;
;
; QY 405 VAOHIAKNPTDALPIPRDPEQVTTFTTEHPWHTQIHRDAFSYGAVGPE 455
; | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; Db 1 VEAHKKKHDPDVLPIPFEPPEQVMIPYTSDFPWHVQVHRDAFSYGDVGPK 51
;
; RESULT 5
; US-09-297-937C-9
; ; Sequence 9, Application US/09297937C
; ; Patent No. 6337199
; ; GENERAL INFORMATION:
; ; APPLICANT: YUM, Do Young
; ; APPLICANT: PAN, Jae Gu
; ; TITLE OF INVENTION: Membrane-Bound Gluconate Dehydrogenase, Gene Sequence
; ; TITLE OF INVENTION: Encoding the Same and Production of 2-Keto-D-Gluconate
; ; TITLE OF INVENTION: Using Transformed Recombinant E. Coli
; ; FILE REFERENCE: P66159US0
; ; CURRENT APPLICATION NUMBER: US/09/297,937C
; ; CURRENT FILING DATE: 1999-05-11
; ; PRIOR APPLICATION NUMBER: PCT/KR98/00296
; ; PRIOR FILING DATE: 1998-09-25
; ; PRIOR APPLICATION NUMBER: KR 97-48802
; ; PRIOR FILING DATE: 1997-09-25
; ; NUMBER OF SEQ ID NOS: 13
; ; SOFTWARE: PatentIn Ver. 2.1
; ; SEQ ID NO 9
; ; LENGTH: 615
; ; TYPE: PRT

```

MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD 97
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,731
FILING DATE: 13-FEB-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 29275/1997
FILING DATE: 13-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: KENNETH H. SONNENFELD
REGISTRATION NUMBER: 33,385
REFERENCE/DOCKET NUMBER: 3479-4000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 751-6840
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 58
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-023-731-11

Query Match 4.0%; Score 130.5; DB 4; Length 58;
Best Local Similarity 44.4%; Pred. No. 3.1e-06;
Matches 28; Conservative 11; Mismatches 13; Indels 11; Gaps 2;

QY 333 YVLASGAINPQILYNSGFGSLQVTPRNDLIPNLGRYITEQPMATCQIVLRQEFVDSVR 392
DB 2 FVACGAVCTPQILWS-----NIRPY-----ALGRYLSEQMTQIVLRGIYDATA 50
QY 393 DDP 395
DB 51 TDP 53

RESULT 8
US-09-265-108-2
; Sequence 2, Application US/09265108
; Patent No. 6033891
; GENERAL INFORMATION:
; APPLICANT: Golightly, Elizabeth
; APPLICANT: Brown, Kimberly
; TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides
; FILE REFERENCE: 5850.000-US
; CURRENT APPLICATION NUMBER: US/09/265,108
; CURRENT FILING DATE: 1999-03-09
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 785
; TYPE: PRT
; ORGANISM: Humicola
US-09-265-108-2

Query Match 3.6%; Score 118.5; DB 3; Length 785;
Best Local Similarity 19.5%; Pred. No. 0.0055;
Matches 126; Conservative 93; Mismatches 257; Indels 171; Gaps 30;
QY 26 AIHETGNDGVDFIAGSGPIGATYAKLCVEAGLRVYVVEIGAADSFYAVNAEEGTAVP- 84
DB 242 APEETY-----DYIVVGAGAGGIPVADKLSEAGHKVLLIEKGPP-----STGRWQGTMKPE 292
QY 85 YVPGYHKKNEIEFQKIDIRFVNVIKALQOVSVFVRNQVNPVLPDGMASAPPGSSAISNG 144
DB 293 WLEG-----TDLTRF--DVPGLCNIWVDSAGIACITDQMGAGCVLGGGTAVNAG 340

QY 145 K--NPHOREFENLSAEAVTRGVGGMSTHTWCSTPRHHPMESLPGIGRKLNSDPAED-- 200
DB 341 LWWKPIDLDWDENFPE-----GWSQDLAAATERV---FERIPGTWHPSDMGKLYRDEG 391
QY 201 -----DKWNELYSEARLIGTSTKEFDESIRHTLVLSRQDAYKQRIFRPLPL 251
DB 392 YKVLSSGLAESGWKEV-----VANEVPEKNKRTFAHTHEMFAGGE-----RNGPL 436
QY 252 ACHRLANAPYVVEVHSAENLFHSIYNDKOKKLTLLTNHRCRLALATGGYEKIGAAEV 311
DB 437 ATY-----LVSADAREN-----FSLWNTAVRRAVETGG---KVTGVEL 472
QY 312 RNLATRNPFSSQDLSYIMAKVYVVLASGAIGNPOILYNSGFGSLQVTPRNDLIPNLGRYI 371
DB 473 ECL--TDGYSYGIVKLNEGGSVIFSAGAFSAKLLFRSGIG----- 511
QY 372 TEQPMATCQIVLR-----QEFVDSVRDDPYGLPMWKEAVAQIAKN-PTDAL-----PIPFR 422
DB 512 ---PEDQLRVVASSKDGEDFIDE-KD-----WIKLPVGYNLIDHLNTDLILTHPDVVVF 561
QY 423 DPEPQVTPTEEPHWHQIHRDAFSYGA--VGP-----EYDSRVIVDL 464
DB 562 DFYEAWTTPIEADKQIYLEQRSIGILAAQAPNIGPMWEEQVTPSGITRQFQWTAARVSGDS 621
QY 465 RMFGATDPEANLLVFQNDVQDCYS-----MPQPTFRYRPSTASNVRAKMMADM 514
DB 622 RTINSSHAMTSLQYLGKGVSVSRGRATITQGLVTVVAEHPYLNHAGDKAEVIOGKNLIES 681
QY 515 CEVASNLGGVLPSP-----QFMDPGL-----ALHLAGTTRIGFDK-----ATTVADN 558
DB 682 LNVIPNITWLP--PPGSTVEEYVDSLLVSASARRSNHMMGTAKLGTDDGRYGTGSVVDL 739
QY 559 NSLVWFDFANLYVAGNCTIRGTGCE-NPTLTSMCHAKSARSINTLK 604
DB 740 DTKVYCTDNLVV-DASIPFGMSTGNPSAMIVIAAEQAERIILKLR 785

RESULT 9
US-09-479-264-2
; Sequence 2, Application US/09479264
; Patent No. 6280976
; GENERAL INFORMATION:
; APPLICANT: Elizabeth J. Golightly
; APPLICANT: Kimberly M. Brown
; TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides
; FILE REFERENCE: 5850.200-US
; CURRENT APPLICATION NUMBER: US/09/479,264
; CURRENT FILING DATE: 2000-01-05
; EARLIER APPLICATION NUMBER: 09/265,108
; EARLIER FILING DATE: 1999-03-09
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 785
; TYPE: PRT
; ORGANISM: Humicola
US-09-479-264-2
Query Match 3.6%; Score 118.5; DB 4; Length 785;
Best Local Similarity 19.5%; Pred. No. 0.0055;
Matches 126; Conservative 93; Mismatches 257; Indels 171; Gaps 30;
QY 26 AIHETGNDGVDFIAGSGPIGATYAKLCVEAGLRVYVVEIGAADSFYAVNAEEGTAVP- 84
DB 242 APEETY-----DYIVVGAGAGGIPVADKLSEAGHKVLLIEKGPP-----STGRWQGTMKPE 292
QY 85 YVPGYHKKNEIEFQKIDIRFVNVIKALQOVSVFVRNQVNPVLPDGMASAPPGSSAISNG 144
DB 293 WLEG-----TDLTRF--DVPGLCNIWVDSAGIACITDQMGAGCVLGGGTAVNAG 340
QY 145 K--NPHOREFENLSAEAVTRGVGGMSTHTWCSTPRHHPMESLPGIGRKLNSDPAED-- 200

```

; INFORMATION FOR SEQ ID NO: 12:
;
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 1876 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;
; MOLECULE TYPE: protein
; US-08-609-049A-12

```

| | | |
|-----------------------|------------------------|---------------------|
| Best Local Similarity | 22.0%; Pred. No. 0.72; | |
| Matches | 85; Conservative | 53; Mismatches 170; |
| | | Indels 79; Gaps 21; |

| | | |
|-----|----|--|
| 83 | QY | VPTVGVYHKKNETEFOKDIRFVNVIKGALQOVSVPRNONVPTLDPGA-----WSAP--P 136 |
| 184 | Db | VYQPAQAQQORPLN--SEELQRLYSM---PAQMAVVPVPOPNATMYPGVAVTPTATPVP 239 |
| 137 | QY | GSSAISNGKNGPHOREFENLSAEAVTRGVGGMSTHWCSTPRTHP--PWSISLPGIG--RPKL 193 |
| 240 | Db | GSAAFVPPQYPAO-----GYGFGAYTHMDLRPQSQAPAPQQTAPTSTSHHSQP 288 |
| 194 | QY | SNDPAPBDDKEWNELYSEAEH---LIGTSTKEPDESIRHTLVLSIQDA-----YKDRQRI 245 |
| 289 | Db | SNHSTSSPAEANGVAFPARQVPTVGVSSSHSTGNNGHSSVPRRGNDLIDLNHEDYSRV 348 |
| 246 | QY | -----FRPIPLACHRUKNAPEYVEWHSAENLPHSIYNDOKKKLFTLLNHRCRTUALG 300 |
| 349 | Db | SVLEAFEDPLLNDNTGNDTASDTSYTYAAYDPDFLYSGD-----AATQYSDPMYEAVN 401 |
| 301 | QY | GYEKKIGIAAEVR--NLATR-----NPSSQLDSYIMA-----KVYVLSAGATGNP----- 343 |
| 402 | Db | RWDKTVATVSPNVGLTGWKQDFLSQPSFTSSSQGVAPPESUKLAENGSGTISPPPLPP 461 |
| 344 | QY | --QILYNSGFGSLQVIT--PRNDSLIPNLGRYITEQPMAFCQIVLRQ-----EFYDSVR 392 |
| 462 | Db | RNQCYESNOAAMPVRSPQSSVLTID--SYTSSIP--ANVYLDRRKCTCTRLVELISDQR 516 |
| 393 | QY | -DDPYGLPWKKEAVAQHIANKPTDALP 418 |
| 517 | Db | TDDPELLEFYF--HMVKEVRARYPHDDAP 542 |

; MOLECULE TYPE

| | | | | | | | | | |
|---------|-----|--------------|-----|------------|------|---------|-----|------|-----|
| Matches | 85; | Conservative | 53; | Mismatches | 170; | Indels. | 79; | Gaps | 21; |
|---------|-----|--------------|-----|------------|------|---------|-----|------|-----|

| | | |
|----|-----|--|
| QY | 83 | VPTVPGVHHKNETEFOKDIDRVNVTKGALQQVSVPRNQNVPTLPDGA-----WSAP--P 136 : : : : : : : : : : : : : : : |
| Db | 184 | VPYQAAQQOORPLN-SEELQRLYSM--PAQMAVVPPVPONMYTYPGAVVTYTATIPV 239 : : : : : : : : : : : : : : : |
| QY | 137 | GSSAISNGKNPHOREFENLSAEAVTRGVGGMSTHWTCSPRIHP-PWESLPGIG--RPKL 193 : : : : : : : : : : : : : : : |
| Db | 240 | GSAAFMPQPQPAO-----GYGFQAYTHMDLRPQSOPAQPOQTAPTTSHHHSQP 288 : : : : : : : : : : : : : : : |
| QY | 194 | SNDPAEDDKEWNELYSABER---LIGTSTKEFDESIRHTLVRLSLQDA----YKDQRRI 245 : : : : : : : : : : : : : : : |
| Db | 289 | SNHSTSAPBAENGVAFPARRQVPSTVGVSSSHGTGNHGSHVPRRGNDLIDLNHEDYSRV 348 : : : : : : : : : : : : : : : |
| QY | 246 | -----FRPLPLACHRUKNAPEYVEWHASBNLFHSIYNDDQKKLFTLLNNHCRTUALTG 300 : : : : : : : : : : : : : : : |
| Db | 349 | SVLEAFPDLNLNDNTGNDTASDTSYYAEYDPDFFLYSGD-----AATQYSDPMYEAVN 401 : : : : : : : : : : : : : : : |
| QY | 301 | GYEKKKGAAEVR-NLAATR-----NPSSQLDSYIMA-----KVYVLASGATGNP----- 343 : : : : : : : : : : : : : : : |
| Db | 402 | RWDKTATVTVSPNVGLTWGRODFLSQPSTSSSQYGVAPEESKUAENGSGTISPPPLPP 461 : : : : : : : : : : : : : : : |
| QY | 344 | --QILYNSGFSGIQVT-PRNDSLIPNLGRYTEBPMAFCQIVLRO-----EFVDYSVR 392 : : : : : : : : : : : : : : : |

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2002, 11:03:12 ; Search time 34.15 seconds
(without alignments)
2010.062 Million cell updates/sec

Title: US-09-856-327-2

Perfect score: 3284

Sequence: 1 MSLSTEQLRDRPSRMQING.....IINTLKGGTGDGKNGEHRNL 618

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------------------|
| 1 | 3284 | 100.0 | 618 | 22 | Lyophyllum shimeji |
| 2 | 1232.5 | 37.5 | 623 | 20 | Coriolus versicolor |
| 3 | 1232 | 37.5 | 633 | 21 | Pleurotus cornucop |
| 4 | 1218.5 | 37.1 | 623 | 17 | Pleurotus cornucop |
| 5 | 1210 | 36.8 | 622 | 22 | Pyrenopeziza triseptata |
| 6 | 1091 | 33.2 | 566 | 20 | Trametes hirsuta p |
| 7 | 1044 | 31.8 | 564 | 21 | An antitumor prot |
| 8 | 1044 | 31.8 | 564 | 21 | T. matsutake pyran |
| 9 | 200 | 6.1 | 544 | 20 | Trichoderma derm |
| 10 | 192 | 5.8 | 51 | 20 | G. oxydans D-sorbi |
| 11 | 191.5 | 5.8 | 615 | 20 | Peptide derived fr |
| | | | | | Gluconate dehydrog |

| | | | | | |
|----|-------|-----|------|----|---------|
| 12 | 177 | 5.4 | 67 | 21 | AA10459 |
| 13 | 177 | 5.4 | 67 | 21 | AA10459 |
| 14 | 168 | 5.1 | 34 | 22 | AA10459 |
| 15 | 162 | 4.9 | 30 | 22 | AA10459 |
| 16 | 159.5 | 4.9 | 657 | 21 | AA10459 |
| 17 | 159.5 | 4.9 | 675 | 21 | AA10459 |
| 18 | 159.5 | 4.9 | 748 | 21 | AA10459 |
| 19 | 142 | 4.3 | 29 | 22 | AA10459 |
| 20 | 130.5 | 4.0 | 58 | 20 | AA10459 |
| 21 | 131 | 3.7 | 24 | 22 | AA10459 |
| 22 | 121 | 3.7 | 704 | 20 | AA10459 |
| 23 | 120.5 | 3.7 | 546 | 20 | AA10459 |
| 24 | 119 | 3.6 | 500 | 22 | AA10459 |
| 25 | 119 | 3.6 | 545 | 21 | AA10459 |
| 26 | 119 | 3.6 | 727 | 21 | AA10459 |
| 27 | 119 | 3.6 | 746 | 21 | AA10459 |
| 28 | 119 | 3.6 | 2639 | 22 | AA10459 |
| 29 | 118.5 | 3.6 | 785 | 21 | AA10459 |
| 30 | 118.5 | 3.6 | 785 | 22 | AA10459 |
| 31 | 116.5 | 3.5 | 698 | 20 | AA10459 |
| 32 | 115 | 3.5 | 50 | 20 | AA10459 |
| 33 | 110 | 3.3 | 551 | 21 | AA10459 |
| 34 | 108 | 3.3 | 35 | 20 | AA10459 |
| 35 | 107.5 | 3.3 | 964 | 22 | AA10459 |
| 36 | 106.5 | 3.2 | 778 | 20 | AA10459 |
| 37 | 106 | 3.2 | 572 | 22 | AA10459 |
| 38 | 106 | 3.2 | 1356 | 22 | AA10459 |
| 39 | 105 | 3.2 | 609 | 22 | AA10459 |
| 40 | 105 | 3.2 | 833 | 22 | AA10459 |
| 41 | 104.5 | 3.2 | 1876 | 18 | AA10459 |
| 42 | 104.5 | 3.2 | 1876 | 22 | AA10459 |
| 43 | 104.5 | 3.2 | 1876 | 22 | AA10459 |
| 44 | 103 | 3.1 | 437 | 22 | AA10459 |
| 45 | 103 | 3.1 | 1686 | 19 | AA10459 |

ALIGNMENTS

RESULT 1

AA10459 standard; Protein; 618 AA.

AC AAB97035;

XX 20-JUL-2001 (first entry)

XX Lyophyllum shimeji antibacterial protein.

XX Fungus; antibacterial; antibiotic; plant pathogen; bacterial infection;
XX Pyricularia oryzae; Rhizoctonia solani; rice pathogen.

OS Lyophyllum shimeji.

XX WO200121657-A1.

XX 29-MAR-2001.

XX 20-SEP-2000; 2000WO-JP06404.

XX 21-SEP-1999; 99JP-0267238.

XX (NISR) JAPAN TOBACCO INC.
XX (NOR) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.

XX Takakura Y, Kuwata S, Inoue Y;

XX WPI: 2001-281598/29.

XX N-PSDB; AAF99980.

XX Antibacterial protein and encoded gene isolated from Lyophyllum
XX shimeji, with activity against plant pathogenic bacteria, applicable in
XX agriculture e.g. rice cultivation at low concentration, produced at low

T. matsutake pyran
N-terminal fragmen
Lyophyllum shimeji
Lyophyllum shimeji
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Lyophyllum shimeji
Peptide derived fr
Lyophyllum shimeji
Candida tropicalis
Choline oxidase (C
A. tumefaciens 1,5
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Novel human diagno
Humicola insolens
H. insolens DSM 18
Candida cloacae fa
Peptide derived fr
Streptomyces globi
Peptide derived fr
Amino acid sequenc
Chlamydia pneumoni
S. epidermidis ope
Drosophila melanog
Propionibacterium
Drosophila melanog
Phosphatidyl inosi
Drosophila melanog
Drosophila melanog
Novel human diagno
Human class II PI3

QY 334 VLASGAIGNPOILYNSGSGQLVTPRND-----SLIPNLGRYTEOPMAFCQIVLRQEF 387
 Db 317 ILTAGVHNAQLIAASFGQI---GRDPKAPLPSLLIPYIGTHITEGLVFCQVMSTEL 373
 QY 388 VDSVRDPP--YGLP-----WKKEAVAQHIKAKNPTDALPIPPFRDPE 425
 Db 374 INSVTADMTIVGKPGHPDYSVITYPGPNPNKHPCDWNEKVKKHHMDHQEDPLPIPEDPE 433
 QY 426 PQTTPTEBHPWHTQIHRDAFSYGAVGPEVDSRVIVDLRWFGATDPEANNLLVFQNDVQ 485
 Db 434 PQVCTLLGATHPWHTQIHRDAFSYGAVGQSIDSLIVDWRFFGTEPKEENKIWSDKIT 493
 QY 486 DGYSHMPQTFRYRSTASNVRAKRMADMCEVASNLGGYLPTSPQFMDPGLALHLAGTT 545
 Db 494 DAYNLKPTDFR--FPGREAEDMMDMCEVMSAKIGGFIpgsyppqfmgglvhlhgth 551
 QY 546 RIGDKAT--TVADNNSLVWDPANLYVAGNTIRTFGENPTLTSMCHAKSARSIIINTL 603
 Db 552 RMGFDEKADCCVDTDSRVFGKFLIGGCGNIPTAYAAANPTLTAMSLAKSCEYI---K 608
 QY 604 KGGTDGKNTGEHRN 617
 Db 609 KNFSPNPVKHHN 622

RESULT 6

AAW87531

ID AAW87531 standard; Protein; 566 AA.

AC AAW87531;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

QY 374 INSVTADMTIVGKPGHPDYSVITYPGPNPNKHPCDWNEKVKKHHMDHQEDPLPIPEDPE 433
 Db 374 INSVTADMTIVGKPGHPDYSVITYPGPNPNKHPCDWNEKVKKHHMDHQEDPLPIPEDPE 433
 QY 426 PQTTPTEBHPWHTQIHRDAFSYGAVGPEVDSRVIVDLRWFGATDPEANNLLVFQNDVQ 485
 Db 434 PQVCTLLGATHPWHTQIHRDAFSYGAVGQSIDSLIVDWRFFGTEPKEENKIWSDKIT 493
 QY 486 DGYSHMPQTFRYRSTASNVRAKRMADMCEVASNLGGYLPTSPQFMDPGLALHLAGTT 545
 Db 494 DAYNLKPTDFR--FPGREAEDMMDMCEVMSAKIGGFIpgsyppqfmgglvhlhgth 551
 QY 546 RIGDKAT--TVADNNSLVWDPANLYVAGNTIRTFGENPTLTSMCHAKSARSIIINTL 603
 Db 552 RMGFDEKADCCVDTDSRVFGKFLIGGCGNIPTAYAAANPTLTAMSLAKSCEYI---K 608
 QY 604 KGGTDGKNTGEHRN 617
 Db 609 KNFSPNPVKHHN 622

RESULT 7

AAAB10457

ID AAB10457 standard; Protein; 564 AA.

AC AAB10457;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Db 80 IKFKDKDKFNIINGALQPISSPDSYQCLAVAAW-APPIDPAEGGLVINGNPNDE 138
 QY 151 EFENLSAEAVTRGVGGMSTHTCTPPIRIHPPMESLPFGIGRPKLSNDPAEDDEWELYSE 210
 Db 139 AGLNIPGSAVTRTVGGMATHWTCACPTPHD-----EERVNPNV-DKEGFALLER 187
 QY 211 AERLIGSTKEPDESIRHTLVLSLDQAYKQRQIFRPLPLACHRLKNAPEYVEVHSAEN 270
 Db 188 AKTLNVHSDQYDSIRQIVVKETLQTL-DASRGVTLPLIGVERTDNPYVWTGAD- 245
 QY 271 LFHSIYVNDKOKKLLTLLNHRCTRLALTGGYEKKIGAEVRLNLLATNPSSQLDSYNA 330
 Db 246 ---TVLGDVPSKPRFAIVTETRVKLIYSETNPTQVVAALLRNL-----NTSN--DELVVA 296
 QY 331 KVVYLASGAIGNPOILYNSGSGQLVTPRNDLIPNLGRYTEOPMAFCQIVLRQEFVDS 390
 Db 297 KSFVIAGAVCTPQLWNS-----NIRPY-----ALGYLSEQSMTCQIVLKGIVDA 345
 QY 391 VRDDPYGLPWKEAQAHIKAKNPTDALPIPRDPEQVTTTTEHPHWHQIHRDAFSYG 450
 Db 346 IATDPR----FAAKEVAHKKHPDVIPIPIHEPEQVPMIPYTSDFPVHGVHRDAFAYG 401
 QY 451 AVGPEVDSRVIVDLRWFGATDPEANNLLVF-----ONDVODGYSMPOPTFRYRST 501
 Db 402 DVGPKADPRVVVDLRFKGSDIVENRVTFGNPKLREWAGVTDYGNPQTFVHKRTN 461
 QY 502 ASNVRAKRMADMCEVASNLGGYLPTSPQFMDPGLALHLAGTTTRIGFDKATTVADNNSL 561
 Db 462 ADGDRDQIMNDMNTNVMIMGGYIPGSGYPPQFMAGPLVHITGTTTIGTDDQTSVADPTSK 521
 QY 562 VMDPANLYVAGNGTIRTFGENPTLTSMCHAKSARSIIINTL 603
 Db 522 VHNFNLLVWVGNGCIPDATACTACNPTRTSAYALKGAEEVNVYL 563

RESULT 7

AAAB10457

ID AAB10457 standard; Protein; 564 AA.

AC AAB10457;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Db 80 IKFKDKDKFNIINGALQPISSPDSYQCLAVAAW-APPIDPAEGGLVINGNPNDE 138
 QY 151 EFENLSAEAVTRGVGGMSTHTCTPPIRIHPPMESLPFGIGRPKLSNDPAEDDEWELYSE 210
 Db 139 AGLNIPGSAVTRTVGGMATHWTCACPTPHD-----EERVNPNV-DKEGFALLER 187
 QY 211 AERLIGSTKEPDESIRHTLVLSLDQAYKQRQIFRPLPLACHRLKNAPEYVEVHSAEN 270
 Db 188 AKTLNVHSDQYDSIRQIVVKETLQTL-DASRGVTLPLIGVERTDNPYVWTGAD- 245
 QY 271 LFHSIYVNDKOKKLLTLLNHRCTRLALTGGYEKKIGAEVRLNLLATNPSSQLDSYNA 330
 Db 246 ---TVLGDVPSKPRFAIVTETRVKLIYSETNPTQVVAALLRNL-----NTSN--DELVVA 296
 QY 331 KVVYLASGAIGNPOILYNSGSGQLVTPRNDLIPNLGRYTEOPMAFCQIVLRQEFVDS 390
 Db 297 KSFVIAGAVCTPQLWNS-----NIRPY-----ALGYLSEQSMTCQIVLKGIVDA 345
 QY 391 VRDDPYGLPWKEAQAHIKAKNPTDALPIPRDPEQVTTTTEHPHWHQIHRDAFSYG 450
 Db 346 IATDPR----FAAKEVAHKKHPDVIPIPIHEPEQVPMIPYTSDFPVHGVHRDAFAYG 401
 QY 451 AVGPEVDSRVIVDLRWFGATDPEANNLLVF-----ONDVODGYSMPOPTFRYRST 501
 Db 402 DVGPKADPRVVVDLRFKGSDIVENRVTFGNPKLREWAGVTDYGNPQTFVHKRTN 461
 QY 502 ASNVRAKRMADMCEVASNLGGYLPTSPQFMDPGLALHLAGTTTRIGFDKATTVADNNSL 561
 Db 462 ADGDRDQIMNDMNTNVMIMGGYIPGSGYPPQFMAGPLVHITGTTTIGTDDQTSVADPTSK 521
 QY 562 VMDPANLYVAGNGTIRTFGENPTLTSMCHAKSARSIIINTL 603
 Db 522 VHNFNLLVWVGNGCIPDATACTACNPTRTSAYALKGAEEVNVYL 563

RESULT 7

AAAB10457

ID AAB10457 standard; Protein; 564 AA.

AC AAB10457;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Db 80 IKFKDKDKFNIINGALQPISSPDSYQCLAVAAW-APPIDPAEGGLVINGNPNDE 138
 QY 151 EFENLSAEAVTRGVGGMSTHTCTPPIRIHPPMESLPFGIGRPKLSNDPAEDDEWELYSE 210
 Db 139 AGLNIPGSAVTRTVGGMATHWTCACPTPHD-----EERVNPNV-DKEGFALLER 187
 QY 211 AERLIGSTKEPDESIRHTLVLSLDQAYKQRQIFRPLPLACHRLKNAPEYVEVHSAEN 270
 Db 188 AKTLNVHSDQYDSIRQIVVKETLQTL-DASRGVTLPLIGVERTDNPYVWTGAD- 245
 QY 271 LFHSIYVNDKOKKLLTLLNHRCTRLALTGGYEKKIGAEVRLNLLATNPSSQLDSYNA 330
 Db 246 ---TVLGDVPSKPRFAIVTETRVKLIYSETNPTQVVAALLRNL-----NTSN--DELVVA 296
 QY 331 KVVYLASGAIGNPOILYNSGSGQLVTPRNDLIPNLGRYTEOPMAFCQIVLRQEFVDS 390
 Db 297 KSFVIAGAVCTPQLWNS-----NIRPY-----ALGYLSEQSMTCQIVLKGIVDA 345
 QY 391 VRDDPYGLPWKEAQAHIKAKNPTDALPIPRDPEQVTTTTEHPHWHQIHRDAFSYG 450
 Db 346 IATDPR----FAAKEVAHKKHPDVIPIPIHEPEQVPMIPYTSDFPVHGVHRDAFAYG 401
 QY 451 AVGPEVDSRVIVDLRWFGATDPEANNLLVF-----ONDVODGYSMPOPTFRYRST 501
 Db 402 DVGPKADPRVVVDLRFKGSDIVENRVTFGNPKLREWAGVTDYGNPQTFVHKRTN 461
 QY 502 ASNVRAKRMADMCEVASNLGGYLPTSPQFMDPGLALHLAGTTTRIGFDKATTVADNNSL 561
 Db 462 ADGDRDQIMNDMNTNVMIMGGYIPGSGYPPQFMAGPLVHITGTTTIGTDDQTSVADPTSK 521
 QY 562 VMDPANLYVAGNGTIRTFGENPTLTSMCHAKSARSIIINTL 603
 Db 522 VHNFNLLVWVGNGCIPDATACTACNPTRTSAYALKGAEEVNVYL 563

RESULT 7

AAAB10457

ID AAB10457 standard; Protein; 564 AA.

AC AAB10457;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Db 80 IKFKDKDKFNIINGALQPISSPDSYQCLAVAAW-APPIDPAEGGLVINGNPNDE 138
 QY 151 EFENLSAEAVTRGVGGMSTHTCTPPIRIHPPMESLPFGIGRPKLSNDPAEDDEWELYSE 210
 Db 139 AGLNIPGSAVTRTVGGMATHWTCACPTPHD-----EERVNPNV-DKEGFALLER 187
 QY 211 AERLIGSTKEPDESIRHTLVLSLDQAYKQRQIFRPLPLACHRLKNAPEYVEVHSAEN 270
 Db 188 AKTLNVHSDQYDSIRQIVVKETLQTL-DASRGVTLPLIGVERTDNPYVWTGAD- 245
 QY 271 LFHSIYVNDKOKKLLTLLNHRCTRLALTGGYEKKIGAEVRLNLLATNPSSQLDSYNA 330
 Db 246 ---TVLGDVPSKPRFAIVTETRVKLIYSETNPTQVVAALLRNL-----NTSN--DELVVA 296
 QY 331 KVVYLASGAIGNPOILYNSGSGQLVTPRNDLIPNLGRYTEOPMAFCQIVLRQEFVDS 390
 Db 297 KSFVIAGAVCTPQLWNS-----NIRPY-----ALGYLSEQSMTCQIVLKGIVDA 345
 QY 391 VRDDPYGLPWKEAQAHIKAKNPTDALPIPRDPEQVTTTTEHPHWHQIHRDAFSYG 450
 Db 346 IATDPR----FAAKEVAHKKHPDVIPIPIHEPEQVPMIPYTSDFPVHGVHRDAFAYG 401
 QY 451 AVGPEVDSRVIVDLRWFGATDPEANNLLVF-----ONDVODGYSMPOPTFRYRST 501
 Db 402 DVGPKADPRVVVDLRFKGSDIVENRVTFGNPKLREWAGVTDYGNPQTFVHKRTN 461
 QY 502 ASNVRAKRMADMCEVASNLGGYLPTSPQFMDPGLALHLAGTTTRIGFDKATTVADNNSL 561
 Db 462 ADGDRDQIMNDMNTNVMIMGGYIPGSGYPPQFMAGPLVHITGTTTIGTDDQTSVADPTSK 521
 QY 562 VMDPANLYVAGNGTIRTFGENPTLTSMCHAKSARSIIINTL 603
 Db 522 VHNFNLLVWVGNGCIPDATACTACNPTRTSAYALKGAEEVNVYL 563

RESULT 7

AAAB10457

ID AAB10457 standard; Protein; 564 AA.

AC AAB10457;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Db 169 -----eervnnpv-dkqefdalleraktlllnvhsdqydsirgqvkvketlqqtl-dasrgv 222
 QY 247 RPLPLACHRLKNAPEVVEVHSAENLFSIYNDDKOKKLLTLLTNHRCRTRALATGGYEKKI 306
 Db 223 tllplgvertdnplyvtwgcad-----twlgdvkpsprvltvtrvtfkfvsetnptqv 278
 QY 307 GAAEVRNLLATRNPSOLSQSYIMAKYVILASGAIGNPQILYNSGFSGLQVTPRNDSLIPN 366
 Db 279 vaallnrl-----ntsn--delvvaqsvlacgavctcpqllwns-----nirph-----a 321
 QY 367 LGRIYIEQPMACQIVLRFQFVDSVRDDPYGLPWKEAVQAQIAKNAKPTDALPIPRDPEP 426
 Db 322 lgrlyseqmtfcqilvkrslvdsiatdpr-----faakveahkkhdpdvlpipfhepeg 377
 QY 427 QVTTPTTEHPWHTQIHRDAFSYGAVGPEVDSRVIVDLRFEGATDPEANLLVLF----- 480
 Db 378 qvmlpytsdipwlnvqhr--yatgvgpkradprvvvdlrffksdlveenvrvcfgnpxl 435
 QY 481 ---QNDVQDGYSPQPTFRYRPTSTASNRARMMADMCEVASNLGGLYLPSPQFMDPGL 537
 Db 436 rdweagvtdtygmppqptfhvkrtnadgdrqmmndmtanvanilgylpgsyppqfmapgl 495
 QY 538 ALHLACTTRIGFDKATTVADNNSLVWDFANLYVAGNGTIRTGEGENPTLTSMCHAKSAR 597
 Db 496 aqhitgttrigtddqtsvadptskvhnfdnlwvgngcipdatacnprrtsvayalkgae 555
 QY 598 SIINTL 603
 Db 556 avvsyl 561

RESULT 9

AAV14049
 ID AAY14049 standard; Protein; 544 AA.

XX AAY14049;

XX 15-JUL-1999 (first entry)

DE G. oxydans D-sorbitol dehydrogenase.

XX D-sorbitol dehydrogenase; L-sorbose; 2-keto-L-gulononic acid; precursor;
 KW L-ascorbic acid production.

XX Gluconobacter oxydans.

XX WO9920763-Al.

XX 29-APR-1999.

XX 13-OCT-1998; 98WO-JP04612.

XX 17-OCT-1997; 97JP-0285280.

XX (FUJI) FUJISAWA PHARM CO LTD.

XX Ishii Y, Noguchi Y, Saito Y, Soeda S, Yoshikawa K;

XX WPI; 1999-302741/25.

DR N-PSDB; AAX57909.

XX Gene group for D-sorbitol dehydrogenase, useful for simple
 PT large-scale production of L-sorbose or 2-keto-L-gulononic acid as
 PT precursor for L-ascorbic acid

XX Claim 7; Page 52-54; 83pp; Japanese.

XX This sequence represents the D-sorbitol dehydrogenase of the
 CC invention. Cells transformed with a vector containing DNA encoding
 CC the dehydrogenase can be used to produce L-sorbose or 2-keto-L-gulononic
 CC acid as precursor for simple large-scale L-ascorbic acid production.

XX Sequence 544 AA;

Query Match 6.1%; Score 200; DB 20; Length 544;
 Best Local Similarity 20.2%; Pred. No. 1.7e-09;
 Matches 134; Conservative 76; Mismatches 220; Indels 232; Gaps 27;

QY 37 DVFIAGSGPIGATYAKLCVZEAGLURVYVMEIGAADSEYAVNAEEGTAVYYPGVHKKNEIE 96
 Db 9 dvvlgsgvagasielanelaraglsvivleagp-----rid 43
 QY 97 FOKDIDRF-VNVIKGALQOVSVPRNQNVPTLPDGAWSAPPSSAISNG-----KNPQRE 151
 Db 44 rghilenfrttenkgvqlypppv-----pwamhpbpdqgsgngylhttgpdgaa 92
 QY 152 FENLSAEAVTRGVGGSTHWCTSTPRHP---PMESLPGIGRPKLSNDPAEDDEKWEELY 208
 Db 93 yq-----qgylrvvvggtwnwagcawrylpdsfelhsrygvgr-----dwaikyddlpefy 143
 QY 209 SEARLIGTSTKEFDESIRHTLVLSLQDAYKDRQRIF--RPLPLACHRLKNAPEYVEWH 266
 Db 144 yqaevmmgvagppmd-----vdlgsprshnypmkvpls-----178
 QY 267 SAENLFHSIYNDDKOKKLLFTLLNHR-----CTRLL 298
 Db 179 -----ygadqfrklkhektnyrvvhhepqrntprpydkrptcegnnncmpicpigm 229
 QY 299 TGG-----YEKKIGAAAEVRNLLATRNPSOLSIVM-----AKVYVLASG 338
 Db 230 yngihsvnhaeagariipnavvyrletdasnkvvvpyvnyydpdknshrvgtgkffvvaah 289
 QY 339 AI-----GNPQILYNSG-----FSLQVLT--PRNDSLIPNLGRYITEQP 375
 Db 290 clesakillllsaddkugprgiansddqvgmmdhtgvgqlsfmgndslwpggrgllt---346
 QY 376 MAFQCIQLRQEFVDSVRDDPYGLPWKE--AVAQHIKAKNFTDALPIFRDPEPOVTPFT 433
 Db 347 -----slidsfrdg---pwsergaylvhmvd-----dnqvdfatgl-380
 QY 434 EEHPWHTQIHRDAFSYGAVGPEVDSRVIVD---LRWF---GATDPEANLLVFQNDVQ 485
 Db 381 -----aiakgvvgkeleeqirygsshavrlfshnegiadpd--nrlslskthk 426
 QY 486 DGYSMPOPTFRYR---PSTASNYRARKMMADMCEVASNLG-----GYLPSPPQFMDPGL 537
 Db 427 dvlgiphevyvylpeytkvscdhtkelfkelmalmstgtdpqwkgyfp-----475
 QY 538 ALHLACTTRIGFDKATTVADNNSLVWDFANLYVAGNGTIRTGEGENPTLTSMCHAKSAR 597
 Db 476 qchpsgstimgtdptnsvvdgecrthdhenlfvarsavfsvgtgnitltigalairvaa 535
 QY 598 SI 599
 Db 536 SI 537

RESULT 10

AAW87541
 ID AAW87541 standard; Peptide; 51 AA.

XX AC AAW87541;

XX 26-FEB-1999 (first entry)

XX Peptide derived from an antitumour protein.

XX Antitumour protein; Tricholoma matsutake.

XX Tricholoma matsutake.

XX JP10313876-A.

XX 02-DEC-1998.

DT 11-DEC-2000 (first entry)
 XX T. matsutake pyranose oxidase protein N-terminal fragment #2.
 XX Pyranose oxidase; matsutake mushroom; antibacterial; sugar metabolism;
 KW diabetes mellitus.
 XX Tricholoma matsutake.
 OS JP2000175698-A.
 XX 27-JUN-2000.
 XX 16-DEC-1998; 98JP-0357423.
 PF 16-DEC-1998; 98JP-0357423.
 PR (NLSB) JAPAN TOBACCO INC.
 XX WPI; 2000-478485/42.
 DR A reagent containing pyranose oxidase for the determination of pyranose
 PT Claim 3; Page 19; 23pp; Japanese.
 XX This invention describes a novel reagent containing pyranose oxidase for
 CC the determination of pyranose which can be prepared from a fraction
 CC precipitated from an aqueous extract of matsutake mushroom by ammonium
 CC sulfate precipitation and has an antibacterial activity against at least
 CC Pyricularia oryzae or Thanatephorus cucumeris and has a M.W. of ca.
 CC 20 kD by gel filtration and shows the presence of components of ca.
 CC 50 kD and ca. 15 kD by SDS-PAGE and maintains the above antibacterial
 CC activity by being heated at 60 degrees C for 10 minutes in a neutral
 CC aqueous solution and in which the above antibacterial activity is
 CC inactivated by being heated at 80 degrees C for 10 minutes in a neutral
 CC aqueous solution. The invention also describes a method for the
 CC determination of pyranose in a sample in which the above reagent for the
 CC determination of pyranose is reacted with pyranose in the sample and the
 CC hydrogen peroxide formed is reacted with an enzyme to develop a color, a
 CC method for diagnosing a disease accompanied by abnormality in sugar
 CC metabolism in which the above reagent for the determination of pyranose
 CC is reacted with pyranose in the sample and the hydrogen peroxide formed
 CC is reacted with an enzyme to develop a color, and a kit for pyranose
 CC analysis or the diagnosis of a disease accompanied by abnormality in
 CC sugar metabolism. The pyranose oxidase can be used as a diagnostic agent
 CC for diabetes mellitus. This sequence represents a Tricholoma matsutake
 CC pyranose oxidase protein N-terminal fragment which is described in the
 CC method of the invention.
 XX Sequence 67 AA;
 SQ

Query Match 5.4%; Score 177; DB 21; Length 67;
 Best Local Similarity 57.4%; Pred. No. 6.8e-09;
 Matches 35; Conservative 6; Mismatches 20; Indels 0; Gaps 0;

OY 484 VODGYSMPQPTFRYPSTASNVRRKMMADMCEVASNLGGYLPTSPQFMDPGLALHLAG 543
 DB 7 vcdtygmpqptfhvkrtnadgdrdgrmndmtvnmmlggyipgysppqfmapgvlvhitg 66
 OY 544 T 544
 DB 67 t 67

RESULT 13
 ID AAY81951
 XX AAY81951 standard; peptide; 67 AA.
 AC AAY81951;
 XX 07-JUL-2000 (first entry)
 DT

XX N-terminal fragment of Trichoderma derived antifungal protein.
 DE Antifungal protein; Pyricularia oryzae; Rhizoctonia solani; mushroom;
 XX growth inhibitor; plant pathogenic fungi; antibacterial agent;
 KW N-terminal fragment.
 XX Trichoderma matsutake.
 OS WO200014242-A1.
 XX 16-MAR-2000.
 PD 19-AUG-1999; 99WO-JP04441.
 XX 08-SEP-1998; 98JP-0270606.
 XX (NLSB) JAPAN TOBACCO INC.
 PA (NORQ) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.
 XX Takakura Y, Kuwata S, Ohta S;
 PI WPI; 2000-256990/22.
 DR Mushroom-derived antibacterial protein against plant pathogenic fungi
 XX of rice, with activity and thermal stability, obtainable cheaply on
 PT large scale, useful in agriculture
 PS Claim 2; Page 27; 52pp; Japanese.
 XX This sequence represents an N-terminal fragment of the protein of the
 CC invention. The protein is an antibacterial protein with activity against
 CC at least Pyricularia oryzae and Rhizoctonia solani. It is obtained from a
 CC fraction of an aqueous extract of a mushroom precipitated by the ammonium
 CC sulphate precipitation method. The protein has a molecular weight of
 CC about 210 kD as determined by the gel filtration method, includes
 CC components of about 15 kD and 50 kD in SDS-PAGE, and is stable to
 CC heating in an aqueous neutral solution at 60 degrees C for 10 minutes but
 CC with loss of antibacterial activity after heating in the solution at 80
 CC degrees C for 10 minutes. The protein is used for inhibiting the growth
 CC of plant pathogenic fungi e.g. Pyricularia oryzae and Rhizoctonia solani.
 CC It is useful in treating rice plants, and is applicable in agriculture as
 CC an antibacterial agent. The protein has activity at relatively low
 CC concentrations, and can be produced at low cost on large scale.
 XX Sequence 67 AA;
 SQ

Query Match 5.4%; Score 177; DB 21; Length 67;
 Best Local Similarity 57.4%; Pred. No. 6.8e-09;
 Matches 35; Conservative 6; Mismatches 20; Indels 0; Gaps 0;

OY 484 VODGYSMPQPTFRYPSTASNVRRKMMADMCEVASNLGGYLPTSPQFMDPGLALHLAG 543
 DB 7 vcdtygmpqptfhvkrtnadgdrdgrmndmtvnmmlggyipgysppqfmapgvlvhitg 66
 OY 544 T 544
 DB 67 t 67

RESULT 14
 ID AAB97034
 XX AAB97034 standard; Peptide; 34 AA.
 AC AAB97034;
 XX 20-JUL-2001 (first entry)
 DT Lyophyllum shimeji antibacterial protein fragment #4.
 DE Fungus; antibacterial; antibiotic; plant pathogen; bacterial infection;
 KW Pyricularia oryzae; Rhizoctonia solani; rice pathogen.
 XX

CC Pyricularia oryzae or Thanatephorus cucumeris and has a M.W. of ca.
CC 210 kD by gel filtration and shows the presence of components of ca.
CC 50 kD and ca. 15 kD by SDS-PAGE and maintains the above antibacterial
CC activity by being heated at 60 degrees C for 10 minutes in a neutral
CC aqueous solution and in which the above antibacterial activity is
CC inactivated by being heated at 80 degrees C for 10 minutes in a neutral
CC aqueous solution. The invention also describes a method for the
CC determination of pyranose in a sample in which the above reagent for the
CC hydrogen peroxide formed is reacted with pyranose in the sample and the
CC method for diagnosing a disease accompanied by abnormality in sugar
CC metabolism in which the above reagent for the determination of pyranose
CC is reacted with pyranose in the sample and the hydrogen peroxide formed
CC is reacted with an enzyme to develop a color, and a kit for pyranose
CC analysis of the diagnosis of a disease accompanied by abnormality in
CC sugar metabolism. The pyranose oxidase can be used as a diagnostic agent
CC for diabetes mellitus. This sequence represents the Tricholoma matsutake
CC invention.
XX
SQ Sequence 564 AA;

Query Match 31.8%; Score 1044; DB 21; Length 564;
Best Local Similarity 40.1%; Pred. No. 7.7e-89;
Matches 243; Conservative 94; Mismatches 197; Indels 72; Gaps 17;

QY 17 QINGQIPKNAIHTYNDGV-----DVFAGSGPIGATYAKLCVE--AGLRVWVVEIGAAD 70
DB 9 kindllqrsqgdltssqdeivhytdvfiagsgpiactyarihiidntsttkvymaiegsqd 68
QY 71 SFYAVNAEETAPVYVPGYHKKNEIEFQKIDIRFVNVIKALQOVSPVRNQNVPTLDGP 130
DB 69 n-----pvi-gahhrnsikfqdktdkfvnlingalqpsispsdyqptlava 115
QY 131 AWSAPPGSSA-----ISNGKNPHOREPENLSAEAVTRGVGMSTHTWCSTPRIHPMESLP 186
DB 116 aw-appldpaegqlvimgnnpqeaaglnlpgsavtrtvggmawhwtacptphd----- 168
QY 187 GIGRPLKSNPDPAEDKKNELYSAEERLIGTSYKEFDESIRHILVLSLODAYKDRORIF 246
DB 169 -----eervnnpv-dkqefdalleraktlinvhdsqydsirqvivketlqgtl-dasrgv 222
QY 247 RPLFLACHRLKNAPEYVEVHSAENLFHSIYNDKQKKLFTLLNHRCTRLALTGGEKKI 306
DB 223 ttllplgvertdnpiyvtgag-----tvlgdvpkpsprflvtrtkfivsetnptqv 278
QY 307 GAEEVRLNLTARNPSSOLDYIMAKVTVLASGAIGNPQILYNSGFSGLQVTPRNDSLPIN 366
DB 279 vaallrnl-----ntsn--delvvaqsfviacgavctpqilwns-----nirph-----a 321
QY 367 LGRVITEOPNAFCQIVLROEFVDSRDDPYGLPWKEAVAQHTAKNPTDALPIFRDPEP 426
DB 322 lgrviseqsmfqcqvlkrsivdsiaedpr-----faakveahkkhdpdvlpihphepe 377
QY 427 QVTTPTTEHPHPTQIHRDAFYSYGAQPEVDSRVIVDLRWFGADPEANLLVF----- 480
DB 378 qmlyptdsfphvqhvr--yafgdvqpkadprvvdlrffgksdlveearvtfgnpkl 435
QY 481 ---QNDVQDSYMPQPTFRYRPTASTASVNRKARKMADCEVASNIGGLYPTSPPOFMDPGL 537
DB 436 rdweagvtdtygmppqtfhvktrnagdrdrqnmndntnvanllgylpgsyqfmapgl 495
QY 538 ALHLAGTTRIGFDKATVADNNSLWDFANLYVAGNCTIRTGGENPTLSMCHAKISAR 597
DB 496 aqhtgttrigtddtsvadqtskvhnfnlhwggngcipdatacnprrtsvayalkgae 555
QY 598 SIINTL 603
DB 556 avvsyl 561

RESULT 8

AAV81952
ID AAY81952 standard; Protein; 564 AA.
XX
AC AAY81952;
XX
DT 07-JUL-2000 (first entry)
XX
DE Trichoderma derived antifungal protein sequence.
XX
KW Antifungal protein; Pyricularia oryzae; Rhizoctonia solani; mushroom;
KW growth inhibitor; plant pathogenic fungi; antibacterial agent;
KW N-terminal fragment.
XX
OS Trichoderma matsutake.
XX
PN WO200014242-A1.
XX
PD 16-MAR-2000.
XX
PF 19-AUG-1999; 99WO-JP04441.
XX
PR 08-SEP-1998; 98JP-0270606.
XX
PA (NISB) JAPAN TOBACCO INC.
PA (NORQ) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.
XX
PI Takakura Y, Kuwata S, Ohta S;
XX
DR WPI; 2000-256990/22.
DR N-PSDB; AAA07403.
XX
PT Mushroom-derived antibacterial protein against plant pathogenic fungi
PT of rice, with activity and thermal stability, obtainable cheaply on
PT large scale, useful in agriculture
XX
PS Claim 3; Page 41-45; 52pp; Japanese.
XX
CC This sequence is the Trichoderma matsutake antibacterial protein of
CC the invention. The protein has activity against at least Pyricularia
CC oryzae and Rhizoctonia solani, and is obtained from a fraction of an
CC aqueous extract of a mushroom precipitated by the ammonium sulphate
CC precipitation method. The protein has a molecular weight of about 210 kD
CC as determined by the gel filtration method, includes components of about
CC 15 kD and 50 kD in SDS-PAGE, and is stable to heating in an aqueous
CC neutral solution at 60 degrees C for 10 minutes but with loss of
CC antibacterial activity after heating in the solution at 80 degrees C for
CC 10 minutes. The protein is used for inhibiting the growth of plant
CC pathogenic fungi e.g. Pyricularia oryzae and Rhizoctonia solani.
CC It is useful in treating rice plants, and is applicable in agriculture as
CC an antibacterial agent. The protein has activity at relatively low
CC concentrations, and can be produced at low cost on large scale.
XX
SQ Sequence 564 AA;

Query Match 31.8%; Score 1044; DB 21; Length 564;
Best Local Similarity 40.1%; Pred. No. 7.7e-89;
Matches 243; Conservative 94; Mismatches 197; Indels 72; Gaps 17;

QY 17 QINGQIPKNAIHTYNDGV-----DVFAGSGPIGATYAKLCVE--AGLRVWVVEIGAAD 70
DB 9 kindllqrsqgdltssqdeivhytdvfiagsgpiactyarihiidntsttkvymaiegsqd 68
QY 71 SFYAVNAEETAPVYVPGYHKKNEIEFQKIDIRFVNVIKALQOVSPVRNQNVPTLDGP 130
DB 69 n-----pvi-gahhrnsikfqdktdkfvnlingalqpsispsdyqptlava 115
QY 131 AWSAPPGSSA-----ISNGKNPHOREPENLSAEAVTRGVGMSTHTWCSTPRIHPMESLP 186
DB 116 aw-appldpaegqlvimgnnpqeaaglnlpgsavtrtvggmawhwtacptphd----- 168
QY 187 GIGRPLKSNPDPAEDKKNELYSAEERLIGTSYKEFDESIRHILVLSLODAYKDRORIF 246

[illegible]

RESULT 5
 AAB48832
 ID AAB48832 standard; Protein; 622 AA.
 XX
 AC AAB48832;
 XX

PT cost on large scale
XX Claim 3; Page 43-45; 52pp; Japanese.
XX
XX The present sequence is an antibacterial protein from the fungus
CC Lyophyllum shimeji. The protein was obtained from a fraction prepared
CC by extracting Lyophyllum shimeji with water and subjecting the extract
CC to ammonium sulphate precipitation. The protein inhibits the growth
CC of the plant pathogenic bacteria Pyricularia oryzae and Rhizoctonia
CC solani at a relatively low concentration. P. oryzae and R. solani are
CC causative of the two major diseases of rice. The protein contains
CC components of 70 kDa and 65 kDa, as determined by SDS-PAGE.
CC The antibacterial protein can be produced at low cost on a large scale.
XX
XX Sequence 618 AA;

Query Match 100.0%; Score 3284; DB 22; Length 618;
Best Local Similarity 100.0%; Pred. No. 2.1e-299;
Matches 618; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSLTEQMLRDYPRSMOINGOIPKNAHETGNDGVDFIAGSGPIGATYAKLCVEAGLR 60
DB 1 mslstegmlrdyprsmginqipknahtyngdgvdfiagsgpigtatyaklcveaglr 60
QY 61 VVMVEIGAADSFYAVNAEETAVPVYVPGYHKKNEIEFQKIDIDRFVNVKIGALQOVSPVR 120
DB 61 vvmveigaadsfyavnaeegtavpvyvpyghkneiefqkldidrfvnvkgalqvsvpr 120
QY 121 NQNVPTLDPGAWSAPPSSAISNGKNPHOREFENLSAEAVTRGVGMSHTWCSTPRIHP 180
DB 121 nqnvtldpgawsappssalsngknpqhorefensaeavtrvgvgmshwtcstprihp 180
QY 181 PMESLPGIGRPKLNDAEDDKENNELYSEARLIGSTKEFDESIRTLVLSLDAYK 240
DB 181 pmeslpgigrpklndaeddkenelysearligstkefdesirtlvlslldayk 240
QY 241 DRQIFRPLPLACHRLKNAPYVWVHSAENLFHSIYNDKOKKFLTLNHRCTRRLALTG 300
DB 241 drqifrpplachrlknapeyvwvhsaenlfhsynddkkfltltnhrctrllaltg 300
QY 301 GYEKKGAAEVRNLLATRNPSQSDSYIMAKVYVLSAGAIQNPOILYNSGFGLOVTPRN 360
DB 301 gyeckgaaevrnllatrnpsqsdtsyimakvyvlsagaignpqllynsfgslqvtpn 360
QY 361 DSLIPNLGRYITEOPMAFCQVLROEFVDSVRDDPYGLPWKEAVAQHIANKPTDALPIP 420
DB 361 dslipnlgrityteopmafcqlvroefvdsvrddpyglpwkeavaqhiaknptdalpip 420
QY 421 FRDPEQVPTTPEEHPWHTQIHRDAFSYGAVGPEVDSRVTVDLRFWFGATDPEANNLIVF 480
DB 421 frdpeqvpttteeHPWHTQIHRDAFSYGAVGPEVDSRVTVDLRFWFGATDPEANNLIVF 480
QY 481 QNVODGYSPQPTFRYRSTASNRARKMADCEVASNIGGLYLPSPFPDPLALH 540
DB 481 qnvodgysspqptfryrstasnrarkmadcevasnlgglylpsppfmdplalh 540
QY 541 LAGTRTGFQKATVANNLSLWDFANLYVAGNCTIRTFGENPTLISMCHAISARSII 600
DB 541 lagtrtfgkattvannslslwdfanlyvagncrtirrtfgenptlismchaiksarsii 600
QY 601 NTLKGGTDGKNTGHRNL 618
DB 601 ntlkgtgdgkntgehrnl 618

RESULT 2
ID AAW94308
XX AAW94308 standard; protein; 623 AA.
XX AAW94308;
XX 08-APR-1999 (first entry)
DT

XX Coriolus versicolor pyranose oxidase.
DE Coriolus versicolor; modified; pyranose oxidase; heat stability;
XX pH stability; glucose; glucosone.
XX Coriolus versicolor.
XX JPI1009271-A.
XX 19-JAN-1999.
XX 24-JUN-1997; 97JP-0181865.
XX 24-JUN-1997; 97JP-0181865.
XX (KIKK) KIKKOMAN CORP.
XX WPI; 1999-145891/13.
XX New modified pyranose oxidase gene and protein - useful for
XX efficient recombinant production of pyranose oxidase with excellent
XX Km, heat and pH stability
XX Claim 1; Page 7-9; 10pp; Japanese.
XX The present sequence represents pyranose oxidase derived from Coriolus
XX versicolor, which can be modified by adding, deleting or replacing at
XX least one amino acid (preferably where the amino acid at position 542
XX is replaced) and still retains pyranose oxidase activity. The modified
XX pyranose oxidase with the amino acid at position 542 replaced, has the
XX following properties: (a) Action: It oxidises glucose to glucosone;
XX (b) Substrate specificity: It reacts specifically with glucose and also
XX reacts with galactose, L-sorbose, D-xylitol and 1,5-anhydro-D-glucitol;
XX (c) Stable pH: 3.5 to 11.0 at 50 degrees Celsius for 30 minutes; (d)
XX Optimum pH: Near 6.5; (e) Optimum temperature: Near 55 degrees Celsius;
XX and (f) Temperature stability: Stable up to 55 degrees Celsius. The
XX present invention also describes a method for the preparation of a
XX modified pyranose oxidase. The method efficiently prepares a modified
XX pyranose oxidase with excellent Km value, heat stability and
XX pH stability.
XX Sequence 623 AA;

Query Match 37.58; Score 1232.5; DB 20; Length 623;
Best Local Similarity 45.0%; Pred. No. 1.7e-106;
Matches 273; Conservative 73; Mismatches 176; Indels 85; Gaps 16;
QY 37 DVFIAGSGPIGATYAKLCVEAGLRVWVVEICGAADSFYAVNAEETAVPVYVPGYHKKNEIE 96
DB 48 dvfiagsgpigtatyaklcveaglvwvveicgaadsvfayvnaeegtavpvyvpyghkneie 95
QY 97 FQKIDIRFVNVKIGALQOVSPVRNQNVPVTLDPGAWSPGSSAISNGKNPHOREFENLS 156
DB 96 yqkndirfwnvkgalqvsvprnqnvpvtldpgawsppgssaisngknphorefens 153
QY 157 AEATRVGGVSTHTWCTSPRIHPHPPMESLPGIGRPKLSNDPAE-DDKEWNELYSEARLI 215
DB 154 gqatrvvgvsthtwctsprihphppmeslpgigrpklsndpae-ddkewnelysaeerli 206
QY 216 GTSTKEFDESIRHTLVLSRQDAYKDRQIFRPLPLACHRLKNAPYVWVHSAENLIF-H 273
DB 207 ktgtdqfkesirhnlvlnkiaeeykg-qrdfqqlplaatr--rsptfveassantvdlq 263
QY 274 SIYNDKOKKFLTLNHRCTRRLALTGGYEKKGAAEVRNLLATRNPSQSDSY----- 327
DB 264 nrpntdapnerfnlpavacerv-----vtn-----tsnseieslhldli 304
QY 328 -----IMAKVYVLSAGAIQNPOILYNSGFGSL-QVTPRN-DSLIPNLGRYITEOPMAFC 379
DB 305 sgdrfeikadvitagavhnaqilvnsfgqlgrpdpapppqilpalsgysiteqsivfc 364

Db 341 LMKPIDLDWENFPE-----GWHSDLAATYERV---FERIPGTWHSPMDGKLYRDEG 391
QY 201 -----DKWENLYSEARLIGSTKEFDESIRHTLVLSLQDAYKDRQIRFRLPL 251
Db 392 YKVLSSGLAESGWKEV-----VANEVPEKNEKTEFAHTHEMFAGGE-----RNGPL 436
QY 252 ACHRLKNAPEYVWHAENLFHSIYNDKQKLFLLTNHRCRTRALATGGYKKGAAEV 311
Db 437 ATY-----LVSADAREN-----FSLWNTAVRRVTRTGG---KVTGVEL 472
QY 312 RNLATRNPSOLDSDYIMAKVYVLASSAIGNPOLLYNSGFSGLQVTPRNDLSLPLNGLRYI 371
Db 473 ECL---TDGYSGLVKNEGGVIFSAFGSAKLLFRSGIG-----511
QY 372 TEOPMAFCOIVLR-----QEFVDSVRDDPYGLPWKKAQAQIAKN-PTDAL-----PIPER 422
Db 512 ---PEDQLRVASSKDGEDFIDE-KD-----WIKLPVGNLDHLENTDLILTHPDVVFY 561
QY 423 DPEPQVTTFTTEHPHHTQIHRDAFSYGA--VGP-----EVDNRVIVDL 464
Db 562 DFEYATTTTTEADKQVLEQSRGILAQAFNIGPMMEQVTPSDGITROPQWTRAVEGDS 621
QY 465 RWFGATDPEANLLVFQNDVQDQYS-----MPQPTFRYPSTASNVARKKXADM 514
Db 622 RTNSHAMTSLQYLGKGVSRGRATITQGLVTVAEHPYLLHNAGKEAVIQIKNLIES 681
QY 515 CEVASNLGGLYLTSP-----QPMDFGL-----ALHLAGTTRIGFDK-----ATTVADN 558
Db 682 LNVIPNITWVLP--PGSTVEEYVDSLLVSASARRSNHNMGTAKLGDDGRYGTSVVDL 739
QY 559 NSLWDFANLYVAGNCTIRTFGE--NPTLTSCHAKSARSINTLK 604
Db 740 DTKVYGTNDLFVY-DASIFPGMTGNPSAMIVIAEQEAERILKLRK 785

RESULT 10

US-09-023-731-14
; Sequence 14, Application US/09023731
; Patent No. 6291648

GENERAL INFORMATION:

APPLICANT: Kawamura, Yukio; Morita,
APPLICANT: Akihiro; Izumo, Koji.; Saka, Tomohide.
TITLE OF INVENTION: ANTITUMOR PROTEIN AND
TITLE OF INVENTION: GENE ENCODING SAME
NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD 97
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,731
FILING DATE: 13-FEB-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 29275/1997
FILING DATE: 13-FEB-1997

ATTORNEY/AGENT INFORMATION:

NAME: KENNETH H. SONNENFELD
REGISTRATION NUMBER: 33,285
REFERENCE/DOCKET NUMBER: 3479-4000
TELEPHONE: (212) 751-6849
TELEFAX: (212) 751-6849
TELEX: 421792

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 50
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-023-731-14

Query Match 3.5%; Score 115; DB 4; Length 50;
Best Local Similarity 53.5%; Pred. No. 9,7e-05;
Matches 23; Conservative 5; Mismatches 15; Indels 0; Gaps 0;

QY 484 VQGYSMQPTFRYRSTASNVARKKXADMCEVASNLGGLYLP 526
Db 8 VTDYGMQPTFRYRSTASNVARKKXADMCEVASNLGGLYLP 50

RESULT 11

US-09-023-731-15

; Sequence 15, Application US/09023731
; Patent No. 6291648

GENERAL INFORMATION:

APPLICANT: Kawamura, Yukio; Morita,
APPLICANT: Akihiro; Izumo, Koji.; Saka, Tomohide.
TITLE OF INVENTION: ANTITUMOR PROTEIN AND
TITLE OF INVENTION: GENE ENCODING SAME
NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD 97
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,731
FILING DATE: 13-FEB-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 29275/1997
FILING DATE: 13-FEB-1997

ATTORNEY/AGENT INFORMATION:

NAME: KENNETH H. SONNENFELD
REGISTRATION NUMBER: 33,285
REFERENCE/DOCKET NUMBER: 3479-4000
TELEPHONE: (212) 751-6849
TELEFAX: (212) 751-6849
TELEX: 421792

INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:

LENGTH: 35
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-023-731-15

Query Match 3.3%; Score 108; DB 4; Length 35;
Best Local Similarity 66.7%; Pred. No. 0.00028;
Matches 22; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 528 SPQFMDPGLALHLAGTTRIGFDKATTVADNNS 560
Db 2 SPQFMDPGLALHLAGTTRIGFDKATTVADNNS 34

RESULT 12

US-08-609-049A-12

; ORGANISM: Erwinia cyripedii
US-09-297-937C-9

Query Match 5.88; Score 191.5; DB 4; Length 615;
Best Local Similarity 21.18; Pred. No. 9.5e-11;
Matches 134; Conservative 83; Mismatches 256; Indels 163; Gaps 32;
QY 36 VDFIAGSGPIGATYAKLCVEAGLRVVMVEIGA-----ADSFYAVNAEEGTAVVPGYH 90
DB 29 VDAVVGVGAGATMAKELTEAGLNVVALERGRHDPYDGAIPQSIDELT-----YN 81
QY 91 KKNIEFQKIDRFVNVKIGALQOVSPVRNQNVPTLDPGAWSPGSSAISNKNPHQR 150
DB 82 IRKRL-FQ-DLSKSTVIRHDASQTAVPYR----- 109
QY 151 EFENLSAEAVTRGVGGSTHTWCSTPRHPPMESLPG-----IGRPKLSNDPAEDD--KE 203
DB 110 ---QLAFLPGTGGAGLHWSGVHFRVDPVELNLRSHYEARYKNTIPEGMTIODEGVS 166
QY 204 WNEL---YSEARLIGTSTKEFDESIRHTLVLSLQDAY--KDRQIRFRPLP-----LAC 253
DB 167 YNELEPFDDQAEKVFGTSGSAW--TIKGMIGKEKGNFYAPDRSSDF-PLPAQKRTYSA 223
QY 254 HRLKNAPEYVEWH-----SAENLFHSIYND----- 279
DB 224 QLFQAASVGYHYDMPDPSANTSGPYNTYGAQMPCNFCGCGYACMYSKASPNVNI 283
QY 280 ---KQKLLTLLNHRCTRLALGGYEKKIGA---AEVNLLATRNPSQ--LDSYIM 329
DB 284 LPALQPKPELRNNAVLRVNLGDKKRATGVYLDGQGREVV---OPADLVLSAFQF 340
QY 330 AKVYVLAGAIGNP-QILYNSGFSGLQVTPRNDLSIPLN-GRYTEOPM---AFCQIVLR 384
DB 341 HNVHMLLSGIGQYPNITNEGVVGRNFAYQNI STLKALFDKNTTNPFFIGAGAGYAVD 400
QY 385 QEFVDSVRDDPYGL-----PWWKEAVAQIAKNPTDALPIPRDPPEQVTPFTTEHPWHT 440
DB 401 DFNADNHDGPGYGVGSPFW-----VNOAGTKPVSGLTPKGTN-----WGS 444
QY 441 QIHRDAPSYGAVGEVDSRVIVDLRWFGATDPEANLLVFQNDVQDGYSPM--OPTERYR 498
DB 445 QWK-----AAVDYTHHHSMDAH--GAHOSYRANYLDLDPNKNYVGPQLLRMTEDWQ 496
QY 499 PSTASNV-ARKMMADCEV--ASN---LGGYLPSPQPMDFGL--ALHLAGTTTRIGF 549
DB 497 D---NDIRMAQFVYGMKMKITEAMNPKMIIGG--AKGPGTHFTDTVTYQTHMSGGAIMGE 551
QY 550 DKATTVADNNSLVWDFANLYVAGNGTIRTGEGENPT 585
DB 552 DPKTSAVNRVLSQSWDVPNVFVPGASAFPOGLGYNPT 587

RESULT 6

US-09-297-937C-13
; Sequence 13, Application US/09297937C
; Patent No. 6337199
; GENERAL INFORMATION:
; APPLICANT: YUN, Do Young
; APPLICANT: PAN, Jae Gu
; TITLE OF INVENTION: Membrane-Bound Gluconate Dehydrogenase, Gene Sequence
; TITLE OF INVENTION: Encoding the Same and Production of 2-keto-D-Gluconate
; FILE REFERENCE: P66159050
; CURRENT APPLICATION NUMBER: US/09/297,937C
; PRIOR FILING DATE: 1999-05-11
; PRIOR APPLICATION NUMBER: PCT/KR98/00296
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: KR 97-48802
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13

; LENGTH: 1276
; TYPE: PRT
; ORGANISM: Erwinia cyripedii
US-09-297-937C-13

Query Match 5.88; Score 191.5; DB 4; Length 1276;
Best Local Similarity 21.18; Pred. No. 3.4e-10;
Matches 134; Conservative 83; Mismatches 256; Indels 163; Gaps 32;
QY 36 VDFIAGSGPIGATYAKLCVEAGLRVVMVEIGA-----ADSFYAVNAEEGTAVVPGYH 90
DB 249 VDAVVGVGAGATMAKELTEAGLNVVALERGRHDPYDGAIPQSIDELT-----YN 301
QY 91 KKNIEFQKIDRFVNVKIGALQOVSPVRNQNVPTLDPGAWSPGSSAISNKNPHQR 150
DB 302 IRKRL-FQ-DLSKSTVIRHDASQTAVPYR----- 329
QY 151 EFENLSAEAVTRGVGGSTHTWCSTPRHPPMESLPG-----IGRPKLSNDPAEDD--KE 203
DB 330 ---QLAFLPGTGGAGLHWSGVHFRVDPVELNLRSHYEARYKNTIPEGMTIODEGVS 386
QY 204 WNEL---YSEARLIGTSTKEFDESIRHTLVLSLQDAY--KDRQIRFRPLP-----LAC 253
DB 387 YNELEPFDDQAEKVFGTSGSAW--TIKGMIGKEKGNFYAPDRSSDF-PLPAQKRTYSA 443
QY 254 HRLKNAPEYVEWH-----SAENLFHSIYND----- 279
DB 444 QLFQAASVGYHYDMPDPSANTSGPYNTYGAQMPCNFCGCGYACMYSKASPNVNI 503
QY 280 ---KQKLLTLLNHRCTRLALGGYEKKIGA---AEVNLLATRNPSQ--LDSYIM 329
DB 504 LPALQPKPELRNNAVLRVNLGDKKRATGVYLDGQGREVV---OPADLVLSAFQF 560
QY 330 AKVYVLAGAIGNP-QILYNSGFSGLQVTPRNDLSIPLN-GRYTEOPM---AFCQIVLR 384
DB 561 HNVHMLLSGIGQYPNITNEGVVGRNFAYQNI STLKALFDKNTTNPFFIGAGAGYAVD 620
QY 385 QEFVDSVRDDPYGL-----PWWKEAVAQIAKNPTDALPIPRDPPEQVTPFTTEHPWHT 440
DB 621 DFNADNHDGPGYGVGSPFW-----VNOAGTKPVSGLTPKGTN-----WGS 664
QY 441 QIHRDAPSYGAVGEVDSRVIVDLRWFGATDPEANLLVFQNDVQDGYSPM--OPTERYR 498
DB 665 QWK-----AAVDYTHHHSMDAH--GAHOSYRANYLDLDPNKNYVGPQLLRMTEDWQ 716
QY 499 PSTASNV-ARKMMADCEV--ASN---LGGYLPSPQPMDFGL--ALHLAGTTTRIGF 549
DB 717 D---NDIRMAQFVYGMKMKITEAMNPKMIIGG--AKGPGTHFTDTVTYQTHMSGGAIMGE 771
QY 550 DKATTVADNNSLVWDFANLYVAGNGTIRTGEGENPT 585
DB 772 DPKTSAVNRVLSQSWDVPNVFVPGASAFPOGLGYNPT 807

RESULT 7

US-09-023-731-11
; Sequence 11, Application US/09023731
; Patent No. 6291648
; GENERAL INFORMATION:
; APPLICANT: Kawamura, Yukio; Morita,
; APPLICANT: Akhito; Izumo, Koji.; Saka, Tomohide.
; TITLE OF INVENTION: ANTITUMOR PROTEIN AND
; TITLE OF INVENTION: GENE ENCODING SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:

Query Match 37.5%; Score 1232.5; DB 1; Length 623;
Best Local Similarity 45.0%; Pred. No. 9.8e-119;
Matches 273; Conservative 73; Mismatches 176; Indels 85; Gaps 16;

QY 37 DVFTAGSGPIGATYAKLCVEAGLRVVMVEIGAADSFYAVNAEEGTAVPYVPGYHKKNIE 96
DB 48 DVIVGSGPIGCTYARELVVAGFVAMFEIGEDSGLK-----GAHKKNIVE 95
QY 97 FOKDIDRFVNVKIGALQOVSVPRNQNVPTLDPGAWSPGSSAISNKNPHQREFENLS 156
DB 96 YQKNIDRFVNVQIGLMSVSPVNTLVITLSPYSQWQ--SSFFVRNGSNPEQDPLRLNS 153
QY 157 AEATRVGGGSMTHWTCTPRIHPMESLPGIGRPKLSNDPAE--DDKEWNLSEAEERLI 215
DB 154 GOAVTRVVGGMSTHWTCTATPFREQ-----RPLLKDDODADDAEMDRLYKAESYF 206
QY 216 GTSKFEDESIRHVLVLSLQADAYKDRQIRPRLPLACHRLKNAPEYVEVHSAENLF--H 273
DB 207 KTGTDQPKESIRHNLVNLKLAEEYKG--QRDPQIPLAATR--RSPTFVWESSANTVFDLQ 263
QY 274 SIYNDKOKKLTLLTNHRCRLALTGGYKKGAAEVRNLLATRNPSQLDSY----- 327
DB 264 NRPTDAPNERENLFPVACERV-----VRN-----TSNSEIESLHIDLI 304
QY 328 ----IMAKYVVLASGAIQNPOLYNSGFSGL-QVTPRN--DSLIPNLGRYITEOPMAFC 379
DB 305 SGDFEIKADVFVLTAGAVHNAQLLVNSGFGQLGRPDANPPQLPSLGSYITEOSLVFC 364
QY 380 QIVLRQBFVDSRDDPY--GLP-----WKKEAVAQIAKHNPTDAL 417
DB 365 QTVNSTELIDSVKMDIIRGNPGLGYSVTYPGAETNKNPDWNEKYNHMQHEDPL 424
QY 418 PIPEDPEQVTTPEEHPHWTQIHRDAFSYGAVGPEVDSRVIVDLRFWFGATPEANNL 477
DB 425 PIPEDPEQVTTQFQSHPHWTQIHRDAFSYGAVQOQSDSLRIVDMFRFGRTPEKENK 484
QY 478 LVFONDYQDQSMQPFPRYR--PSTASNVARKMADMCVNASNLGGYLPSTPPQFMDPG 536
DB 485 LWFSDKITDVTNMQPTDFRFPAGRTSKEAEDMTDMCVMSAKIGGLPGLSLPQFMPEG 544
QY 537 LALHAGTTRIGFPAK--TTVADNNSLVWDFANLYVAGNGIRTFGENPTLTSCHAIK 594
DB 545 LVHLGTHRMGFDEQEDKCCVNTDSRVFGKFLGCGGNIPATYAGNPTLTAMSLAIK 604
QY 595 SARSIIN 601
DB 605 SCEYIKN 611

RESULT 2
US-09-305-381-2
; Sequence 2, Application US/09305381
; Patent No. 614865
; GENERAL INFORMATION:
; APPLICANT: Christensen, Soren
; APPLICANT: Lassen, Soren Flensted
; APPLICANT: Schneider, Palle
; TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides
; FILE REFERENCE: 5571.200-US
; CURRENT APPLICATION NUMBER: US/09/305.381
; CURRENT FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: 60/088.724
; EARLIER FILING DATE: 1998-06-10
; EARLIER APPLICATION NUMBER: PA 1998 00774
; EARLIER FILING DATE: 1998-06-08
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 622
; TYPE: PRT

; ORGANISM: Coriulus Versicolor
us-09-305-381-2

Query Match 36.8%; Score 1210; DB 4; Length 622;
Best Local Similarity 44.6%; Pred. No. 2.1e-116;
Matches 274; Conservative 71; Mismatches 197; Indels 72; Gaps 16;

QY 37 DVFTAGSGPIGATYAKLCVEAGLRVVMVEIGAADSFYAVNAEEGTAVPYVPGYHKKNIE 96
DB 48 DVIVGSGPIGCTYARELVVAGFVAMFEIGEDSGLK-----GSHKKNIVE 95
QY 97 FOKDIDRFVNVKIGALQOVSVPRNQNVPTLDPGAWSPGSSAISNKNPHQREFENLS 156
DB 96 YQKNIDRFVNVQIGLMSVSPVNTLVITLSPYSQWQ--SSFFVRNGSNPEQDPLRLNS 153
QY 157 AEATRVGGGSMTHWTCTPRIHPMESLPGIGRPKL--SNDPAEDDKENWNLSEAEERLI 215
DB 154 GOAVTRVVGGMSTHWTCTATPFKEK-----LQRPVLVKNDSKADDAEMDRLYKAESYF 206
QY 216 GTSKFEDESIRHVLVLSLQADAYKDRQIRPRLPLACHRLKNAPEYVEVHSAENLF--H 273
DB 207 KTGTDQPKESIRHNLVNLKLAEEYKG--VRDPQIPLAATR--QSPTFVWESSANTVFDLE 263
QY 274 SIYNDKOKKLTLLTNHRCRLALTGGYKKGAAEVRNLLATRNPSQLDSYIMAKYV 333
DB 264 NRPNKADPKQFENLFPVACTNVRNDRNANSEIVG--LDVRDLHGKKSIT-----IKAKYV 316
QY 334 VLASCAIGNPOLYNSGFSGLQVTPRND-----SLIPNLGRYITEOPMAFCOIVLROEF 387
DB 317 ILTAGAVHNAQLLAASGFGQL---GRDPAPAKPLSLPYLGTHTHTTEQTLVFCQVNSTEL 373
QY 388 VDSVRDDP--YGLP-----WKKEAVAQIAKHNPTDALPIPFDRPE 425
DB 374 INSVTADMTIVGKPGHPDYSVTYTPGNPNKHPDWNEKYNHMQHEDPLPIPFDEPE 433
QY 426 POVTPPTTEHPHWTQIHRDAFSYGAVGPEVDSRVIVDLRFWFGATDPANNLLVFQNDVQ 485
DB 434 POVTTLFOATPHWTQIHRDAFSYGAVQOQSDSLRIVDMFRFGRTPEKENKLFSDKIT 493
QY 486 DQSNMPOQTFYRPSASTASNVARKMADMCVNASNLGGYLPSTPPQFMDPGLALHAGTT 545
DB 494 DAYNLRQTFDEFR--FPGREADEMTDMCVMSAKIGGLPGLSGYPOFMERGLVHLGTH 551
QY 546 RIGEDKAT--TVADNNSLVWDFANLYVAGNGIRTFGENPTLTSCHAIKARSIIINTL 603
DB 552 RMGFDEKADKCCVNTDSRVFGKFLGCGGNIPATYAGNPTLTAMSLAIKSCYI---K 608
QY 604 KGGTDGKNTGEHRN 617
DB 609 KNFEPSNPVKHN 622

RESULT 3
US-09-023-731-1
; Sequence 1, Application US/09023731
; Patent No. 6291648
; GENERAL INFORMATION:
; APPLICANT: Kawamura, Yukio; Morita,
; APPLICANT: Akihito; Izumo, Koji.; Saka, Tomohide.
; TITLE OF INVENTION: ANTI-TUMOR PROTEIN AND
; TITLE OF INVENTION: GENE ENCODING SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE

Db 109 VER-----EDGVSPAPVSYAELEPYAEALFGV-RGRAGDDPTPEPRSPAPMH 158
QY 209 SEA--ERLIGTSKEFDESIRHTLVLSRQDAYKDRQIRPLPLA-----CHRLKN 258
Db 159 APIPEPVIGRAVGPER-----LGLRPFH-----MPSAIDYGPGLCRRCGT 201
QY 259 APYVWHSANLPHSIYNDKOKKLF-----TLTNHRCRTRALALGGYEKKIGAA 309
Db 202 CDAFVCRFDAGK-----DAETRLRLPALRHPNVSLLTGARVRL-IADGDKHIVAV 252
QY 310 EVANLLATRNPPSOLDYSIMAKVYVLASGAIGNPQILYNSGFGSLQVTPRNDLSLIPN--- 366
Db 253 EIER-----AGEIIT-IEAPLFLVSAGAINSALILLRSA-----DEKKNGLA 294
QY 367 -----LGRVITEOPMA-----FCQIVLRQEFVDSVRDDP----- 395
Db 295 NSSGVYGRVILNMHLSGLMLLPFTINDTRFKTMSLNDPFGTGGDEAAAGNVQMLGNI 354
QY 396 -----YGLPWKEAVAQHTAKNPTDALPIPRDPPEQVTPPTTEHPWHTQIHRDAFSY 449
Db 355 QGPMIRAAIWPMPRLANLARSVDFLV-----SED----- 387
QY 450 GAYGPEVDSRVIVDLRFWFGATDEANLLVFQNDVDGYSMPQPTFRYRSTASN-----V 505
Db 388 ---TPKYDSRV-----KPGKNGAE-----LIYRPGDREAHQRFV 419
QY 506 RAKMADMCVANSNGVLYPTSPFMDPGLALHAGTTRIGDKATTVADNNSLWDF 565
Db 420 RHRSLRLKNGFPVVLGSHSGIEAPS-----HOCGTVMRGDDPKREAAALNLCQTYDH 471
QY 566 ANLYVAGNGTIRTFGENPTLTSMAIHK 594
Db 472 PNLYVVDAGFFPSSAALNPALTPVAAQALR 500

RESULT 14
JC6564
cellobiose oxidase (EC 1.1.3.25) precursor - white-rot fungus (Trametes versicolor)
N:Alternate names: cellobiose dehydrogenase
C:Species: Trametes versicolor (white-rot fungus)
C:Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 19-May-2000
C:Accession: JC6564
R:Dumoncaux, T.J.; Bartholomew, K.A.; Charles, T.C.; Moukha, S.M.; Archibald, F.S.
Gene 210, 211-219, 1998
A:Title: Cloning and sequencing of a gene encoding cellobiose dehydrogenase from Trametes
A:Reference number: JC6564; MUID:98248903
A:Accession: JC6564
A:Molecule type: mRNA
A:Residues: 1-768 <DUM>
A:Cross-references: GB:AF029668
A:Experimental source: 52J, AFCC 20869
C:Comment: This is a hemoflavoenzyme that oxidizes cellobiose and reduces iron(III) and
C:Genetics:
A:Gene: cdh
A:Introns: 17/1; 103/2; 140/1; 179/1; 226/3; 281/3; 306/1; 324/2; 534/3; 577/2; 593/3; 6
C:Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-768/Product: cellobiose dehydrogenase #status predicted <MAT>
F:80,128/Binding site: heme iron (Met, His) (axial ligands) #status predicted

Query Match 4.2%; Score 138.5; DB 2; Length 768;
Best Local Similarity 19.2%; Pred. No. 0.032;
Matches 128; Conservative 83; Mismatches 228; Indels 229; Gaps 33;

QY 37 DVFIAGSGPTGATYAKLCVAGLRVNVVEIG-----AASFYAVNAEETAVPY-VPG- 88
Db 231 DYIVVAGPGGLVADRLSAGKVKLLERGGSTATGTYDATWAKSANLTKFDVPLG 290
QY 89 -----YHKKNIEFQKIDRFVNVVNIKALQVSVPRNQNVPTLDPGAWSPGSSATNG 144
Db 291 FEFLTDNPFWCKDNTNFFAGCLLG-----GGTSV-NG 323

QY 145 K---NPHQREFNLSAEAVTRGVGGMSTHTWCSTPRHHPPEMLPGIGRKLKSLNDPAEDD 201
Db 324 ALXYWPNRDFSTAS-----GWPSWSNHQPTDKLQRLPS-----TDHPSADG 368
QY 202 KENNELYSEA--ERLIGTSKEFDESIRHTLVLSRQDAYKDRQIRF-----RP 248
Db 369 QRYLESATVVOQLSQGYSQI-----TINDNPDSDKHVGFSAFDFLNGQRA 417
QY 249 LPLACHRLKNAPYEVWHSANLPHSIYNDKOKKLFLLTNHRCRTRALALGGYEKKIGA 308
Db 418 GPVA-----TYFETALARKNE--VYKDN---VLVTQVIRNGSTILGVRTN-DNTLGP 463
QY 309 AEVRNLLATRNPPSOLDYSIMAKVYVLASGAIGNPQILYNSGFG---LQVTPRNDLSLIP 365
Db 464 DGIVPL-----NPNGRV-----ILSGSGFTPRILFOSGIGPTDMLQTVOSNAQAAA 510
QY 366 NLGRYTEQPMATFCQIVLROEFVDSVRDDPYGLPMMKEAVAQHTAKNPTDALPIPRDPE 425
Db 511 NL-----PP-----QSEWID-----LP-----VGQSVSDNPS----- 532
QY 426 PQVTTPTTEHP-----WHTQIHRDAFSY-----GAVGPEVDSRVIVDLRW 466
Db 533 ---INLVET--HPSIDAYDNWADVMSNPRPADAAQYQLQSRGVLGASPKLNF-----WRA 583
QY 467 FGATDEANLLVFQNDVDGYSMPQPTFRYRPS-----TASN 504
Db 584 YGSGDITR---YAQGTVRPGAASVNTSVAINASEIPTITLYLSNGIQSRIGRIGVDMALN 640
QY 505 VRA-----RKMDADMCVANSNGVLYPTSPFMDPGLALHAGTTRIGDKATTVADNNSLWDF 538
Db 641 AKALVNPWLTVNSDKTVLLQALHDVTSTMKNVPLGTLITPDNTMTLEQYVAAVDPATMCS 700
QY 539 LHLAGTTRIGDKATTVADNNSLWDFANLYVAGNGTIRTFGENPTLTSMAIHKARS 598
Db 701 NHWVGAAMGTSSSTAVVDENAKVFENTDNLFIVDASIIPSLPICNGQVLM-SAAEQAVS 759
QY 599 IINTLKG 606
Db 760 RILALAG 767

RESULT 15
AD3255
l-sorbose dehydrogenase (FAD) (EC 1.1.99.-) [Imported] - Brucella melitensis (strai
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AD3255
R:DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivan
.; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; L
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella mel
A:Reference number: AD3252; PMID:11756688
A:Accession: AD3255
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-494 <KUR>
A:Cross-references: GB:AE008917; PIDN:AAL51207.1; PID:g17981893; GSPDB:GN00190
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI0025
A:Map position: I
C:Keywords: oxidoreductase

Query Match 4.2%; Score 138; DB 2; Length 494;
Best Local Similarity 20.4%; Pred. No. 0.017;
Matches 125; Conservative 67; Mismatches 240; Indels 182; Gaps 23;

QY 37 DVFIAGSGPTGATYAKLCVAGLRVNVVEIGADSFYAVNAEETAVPYVPGYHKKNIE 96
Db 6 DIVIIGSGIGGATMAAGLAAGADILILEAGARLADRPENRDPRAI--FORGFRPKELW 63

Db 341 FGSQYQSNDEVTIRIGYTCVKHGFN-----QVLKGGERGLQVSEVPRNSPDH 395
QY 262 -----YVEWHAENLFHSIYNDKQKFLTLTHNRCTRLAL-----TGGEYKK-----I 306
Db 396 YGGLGCGYGCAGAKNGTDQTLVDVAVENGAVILTGKAERFVLVDNTSSNERKKRCVGV 455
QY 307 GAAEVRNLLATNPSSQDLSYIMAKYVVLASGAIGNPOLYNSGSGLOVTPRNDSLIPN 366
Db 456 FASSVGGKIGK-----FIIEARVTSSAGSLTLPMLSSGLKN-----PN 497
QY 367 LGR-----YITEOPMAFCQIVLRFQFVDSVR--DDPYGLPWNKEAQAQIAKN 412
Db 498 IGRNKLHPVMTWGYFPKDESEFGKTEGGIITSVHMND-----TESGCKAILLEN 550
QY 413 PTDALPIPRDPPEQVTPFTTEEHFWHTQIHRDAFSYGAVGEVDSRVIVDLRFWFGATDP 472
Db 551 PLIG-PASVAGLSPWVS-----GPDLMKRMK-----YGRY-- 580
QY 473 EANNLLVFQNDVQDYSM--POPTFRYRSTASNVARKMMADMCEVAS--NLGGY-- 524
Db 581 --AHLFALVRDLGSEVAMENEVYRTTKDRENLRAGLRQALRVSAAGAIVEGVYRSD 638
QY 525 -----LPTSPQEMDPGLAL-----HLAGTRIGFDKATTVADNN 559
Db 639 GOKMCEALITKAMEEFLDEYDAVGVGKGYWYTFSAHQMGSCRMGVTAEEGALDEN 698
QY 560 SLVWDFANLYVAGNRTIRTFGENPTLTSMCHAIRKSARSIIINTLKGTT 607
Db 699 GESWEAGLFVCDGSLPSAVGNPMITIQSTAYCISSKIVDSLQNT 746

RESULT 10
E98192
probable oxidoreductase chain CJ0415 [imported] - Agrobacterium tumefaciens (strain C58,
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C:Accession: E98192
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; PMID:11743194
A:Accession: E98192
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-561 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK89063.1; PID:g15158859; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L_980
A:Map position: linear chromosome

Query Match 4.5%; Score 148.5; DB 2; Length 561;
Best Local Similarity 19.1%; Pred No. 0.0033;
Matches 135; Conservative 60; Mismatches 222; Indels 289; Gaps 29;

QY 37 DVFIAGSGPIGATYAKLCVAGLRVYVMEIGAASFYAVNAEEGTAVPYVGYHKKNETE 96
Db 7 DAIVVSGISGGWAANKLTQKGLKVLML-----ERGRNIEHVDY----- 46
QY 97 FOKDIDRFVNIKALQVSVPRNVPTLDPGAWSPGSSAISNGKNPHOREFENLS 156
Db 47 -----QNA-----DKEAWDYPHNRNATQEMKAKYVLSRDYL 78
QY 157 AEAVTRG-----VCGMSTHTCTSPRIHPHPPMESLPGIGRPKL 193
Db 79 LEATLGMWADQETPYVEEKRFDFRGRVHVGSRLLMGRQTYRW--SOTDFEANAQGI 136
QY 194 SND---PAEDDKENNELYSEARLICTSTKEPDESIRHTLVLSLQDQYDKQRIFRPLP 250
Db 137 AVDWPIRVEDVSPW---DYVERFAGISG-----REGLDILPDGE-FLPPIP 180

QY 251 LACHRLKNAPEVVEVHSAENLFHSIYNDKQKFLP---TLLNHRCTRALT----- 299
Db 181 L-----NEVEDQVASRL-----KKAFFKGRHLINRSRANITQELPDQERTP 221
QY 300 -----GGY-----EKKIGAAEVRNLLATR-----N 319
Db 222 CQFRNKRCLGCGPFGYFTSQASTLPAAVATGNLTURPFSIVIREILYDKKKARGVEIID 281
QY 320 PSSQDLSYIMAKYVVLASGAIGNPOLYNSGF-----SGLOVTPRNDSLIPNCRYITE-- 373
Db 282 AETNLTYEYADIIIFLNASTLNTSVLMKNSATDVWEGGL-----CSSSGELGHNVMDDH 335
QY 374 -----OPMAFCQIVLRFQFVDSVR--DDP-----YGLP----- 399
Db 336 FRMGATQVDGFEDEFYFKGRPAGE-----YIPFRFNTGDKRKYLRGFGYOGSASR 387
QY 400 --WKEAQAQHTAKNPTDALPIPRDPPEQVTPTEEHFWHTQIHRDAFSYGAVGEV 457
Db 388 SRWEIEALNIGADYKDTL-----TEPGGW--TIGMTAF--GEMLPYHD 428
QY 458 SRVIVDLRFWFGATDPEANLLVFQNDVQDYSMPOTFRYRSTASNVARKMMADMCE- 516
Db 429 NRVKLD-----HDKKDKWGL-----PVLSNNVEMKQNELDMRED 462
QY 517 -----VASNLGGYLPSTPPQFMDPGLALHLAGTRIGFDKATTVADNN 560
Db 463 MYNDVAVEMPEAVGKRVKPSRGTYA-----PCMGIHMGRTARMGRDPKTSVLNGNN 513
QY 561 LWDFANLYVAGNRTIRTFGENPTLTSMCHAIRKSARSIIINTLKG 606
Db 514 QVWDAPNFVTDGACMTSASCNVNLTYNALPARAAEFVSRKKG 559

RESULT 11
AE3094
oxidoreductase Atu4377 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AE3094
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Wc
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McC
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Ka
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AE3094
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-561 <KUR>
A:Cross-references: GB:AE008689; PIDN:AA45171.1; PID:g17742847; GSPDB:GN00187
C:Genetics:
A:Gene: Atu4377
A:Map position: linear chromosome

Query Match 4.5%; Score 148.5; DB 2; Length 561;
Best Local Similarity 19.1%; Pred No. 0.0033;
Matches 135; Conservative 60; Mismatches 222; Indels 289; Gaps 29;

QY 37 DVFIAGSGPIGATYAKLCVAGLRVYVMEIGAASFYAVNAEEGTAVPYVGYHKKNETE 96
Db 7 DAIVVSGISGGWAANKLTQKGLKVLML-----ERGRNIEHVDY----- 46
QY 97 FOKDIDRFVNIKALQVSVPRNVPTLDPGAWSPGSSAISNGKNPHOREFENLS 156
Db 47 -----QNA-----DKEAWDYPHNRNATQEMKAKYVLSRDYL 78
QY 157 AEAVTRG-----VCGMSTHTCTSPRIHPHPPMESLPGIGRPKL 193
Db 79 LEATLGMWADQETPYVEEKRFDFRGRVHVGSRLLMGRQTYRW--SOTDFEANAQGI 136

Db 203 RKAEEELGLHPYDLPAANASGPTWNPYGVOMGPNFCGFCGAGYACTWYSKASPNL--NIL 260
QY 277 NDDKQKQLFTLLNHRCTRRLALGGYEKKIGA-----AEVRNLLATRNPSQOLDSYIMAKV 332
Db 261 PALRQTPLELRANCNVLYKVNLDSDGROATGVTVYVDAQGREIVQP-----AKL 308
QY 333 YVLASAGNAPQILYNSGSLQVTPRNDLSLNLGRYITEQPMACQIVLROEFVDSVR 392
Db 309 VIISAFQFHNVRLLLSGI--GKPYDPTGEGV--VGKNFAIQNMA-----TIKAFQKDVH 361
QY 393 DDYF-----GLPWWKEAVAQIAKNAKPTDALPIPPRDE 425
Db 362 TNPFTVGTGGGVAVDNADNFDHGLPFGVGGSPMW-----VNGAGSKPIGLAVPPGTPS 417
QY 426 PQVTTPTTEHPWHT---QIHRDAFSYGAVPEVDSRVIVDLRWFGATDPEANNLLVFQN 482
Db 418 -----WGSQWKAQVADAYTH-----TVSDAHGSMNTYRDYLDLP 454
QY 483 DVODGYSMP--QPTFRY-----RPSTASNVRAKMMADCEVA-----SNLGGYLPSP 530
Db 455 TYKDATGQPLRLMTFDKNEIRMSRYVTEHMKRIAEAMNPKAISVSVKRFGDHFNTRYV 514
QY 531 QFMDPGLALHLAGTTRIGFDKATTVADNNSLVWDFANLYVAGNGTIRTFGENPTLTSMC 590
Db 515 Q-----TTHLLGGAINGSDPKTSVLNRYLOSNDVHNVFVWAGASAPPGTGYNPTGLVAA 568
QY 591 HAIKSARSI 599
Db 569 LAYWSAKAI 577

RESULT 6
A84260
hypothetical protein Vng1035c [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: A84260
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.;
Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabil
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebbardt, H.; Lowe, T.M.; Li
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483
A:Accession: A84260
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-529 <STO>
A:Cross-references: GB:AE004437; NID:g10580587; PIDN:AAG19445.1; GSPDB:GN00138
C:Genetics:
A:Gene: VNG1035C

Query Match 5.5%; Score 179.5; DB 2: Length 529;
Best Local Similarity 22.1%; Pred. No. 1.3e-05;
Matches 143; Conservative 53; Mismatches 236; Indels 215; Gaps 28;
QY 37 DVFIAGSGPTGATYAKLCVEAGLRVYVVEIGAADSFSYVAVNAEGTAVPYVPGYHKRKEIE 96
Db 10 DVCVIGAGPAGGLIADRLASDGHVVVLEAG--PRDAADR----- 50
QY 97 FOKIDRFVNIKGAQQVSVVRNQNPTLPDPAAS--APPSSAISNGKNPHOREFENL 155
Db 51 -----RMERSIRPAHGPAV-----WEMGPRDAYASTGD-----RYPL 85
QY 156 SBAEATRVGGGSMTHWCTCTPRTHPP---MESLPGIGRPLKSLNDPAEDKKNELYSEAE 212
Db 86 NAARV-KGVGGSTLHQGMVMRLHEQDFLASATGVA-----DWPPDYDTLPKYAAAE 139
QY 213 RLIT---GTSTKEPDESIRHTLVLSRLQDAYKDRQIRFRLPLACHRLKNAPEYVHSAE 269
Db 140 SALVSGASDNFPAPPREQHPQAPPPPSYSD--SLFAD---ACESLGIAI----- 195

QY 270 NLPHSYNDKQKQLFTLLNHRCTRRLALG-----GYE-----K 304
Db 186 ---HSPFN-----ARLSAGRETRACVGYGTCQPCPSGAKYDATVHVDRATDAGAR 234
QY 305 KIGAAEVRNL-----LATNPSSQLDSYIMAKVYVVLASCAIGNPOIL----- 346
Db 235 VIDEAPVORLEHDAAGDRVTGAVYATPDGTTTHROS---ATEFLAAGGIETPRLLLSDS 291
QY 347 --VNSGF---SGLQVTPRNDLSLNLGRYITEQPMACQIVLROEFVDSVRDDPYGLPMW 401
Db 292 DRYPDGLANSSGL-----VGRY-----FMD----- 311
QY 402 KEAVAQIAKNAKPTDALPIPPRDEPQVTTPTTEHPWHTQIHRDAFSYGAV-----G 433
Db 312 -----HLFAGAGTLDDETRONHVGEHT--TESHOY---DRPDGSRGAIKLEFLNYAG 360
QY 454 PEVDSRVIVDLRFGA-----TDPEANLLVFQNDQVQDYSMPQ 493
Db 361 PSFAEMALSDDMDAMCDIRDASTHIAVGGVLEVOQPPRPNRVLHPERTDVHGNVP 420
QY 494 TFRYRPTSTASNVRAKMMADCEVASNLGGYLP--TSPQPFMDPGLALHLAGTTRIGFDKA 552
Db 421 DVWVSLSAVERRTIERANEIOREILTELGADEIETVGP--DTGPAFHHMGTTRMGTDPA 478
QY 553 TVVADNNSLVWDFANLYVAGNGTIRTFGENPTLTSCHAIKSARSI 599
Db 479 ESYVDPRLRTHDLSNLSVASSVFPPTAGAMNPTLTIAALAKAADHI 525

RESULT 7
B75608
GMC oxidoreductase - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: B75608
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.
S.; Smith, H.O.; Venter, J.C.; Lam, P.; Mcdonald, L.; Utterback, T.; Zalewski, C.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1
A:Reference number: A75250; MUID:20036896
A:Accession: B75608
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-722 <WHI>
A:Cross-references: GB:AE001862; GB:AE001825; NID:g6460468; PIDN:AAF12230.1; PID:g6.
A:Experimental source: strain R1
C:Genetics:
A:Gene: DRA0127
A:Map position: 2

Query Match 5.4%; Score 176; DB 2: Length 722;
Best Local Similarity 20.9%; Pred. No. 3.8e-05;
Matches 139; Conservative 77; Mismatches 239; Indels 210; Gaps 33;
QY 14 RSMQINGQIPKNAIHTYTGNDG---VDVFIAGSGPTGATYAKLCVEAGLRVYVVEIGA- 68
Db 186 RNSAITPYVPQ-----DGEVLEADAVVVGSGGGVIAARLAQAGKRVVLEAGGH 236
QY 69 --ADSFYAVNAEGTAVPYVPGYHKRKEIEFQKIDIRFVNIKGAQQVSVVRNQNPT 126
Db 237 YBEAHFGRELAAYQTLTYRGVH-----PTADGNV-T 268
QY 127 LDPAWSAPPSSAISNGKNPHOREFENLSAEAVTRGVGGMSTHWTCTPRIPHPHESLP 186
Db 269 LVAGA-----NLG-----GGSTVWNSV-----PPRDDI- 293
QY 187 GIGRPKLSN-----DPAEDDKKNELYSEAEIRLTSTKEFDESIRHTLVLSRLQDA 238
Db 294 ---RQWASEHGLSDVADPGYD---RHIDAVLERM-GVSEQCSHDNGPHQ---RLVEGA 342
QY 239 YKQRQIRFRLPLACHRLKNAPEYVHSAENLFHSIYND---KOKKLT----- 286

QY 393 DDPYG--LPWKE-----AVAQIAKNPTDALPIPRDPDPQVTPPTTEHPHWHQIH 443
Db 362 DDFGDNFDHSEKELHGMAMISVQLGTRPIQASAPLPAGAP-----TWGAEP- 408
QY 444 RDAESYG-----AVGPEVDSRVIVDLRFEAGTDPANLLVFNQVQDGYSPQPTFRYR 498
Db 409 KKALYNFTRAITVGGQ-----GASLPKNNYLSLOPTYKDAFGMPLRLRTYN 456
QY 499 PSTASNVRARKMAD-MCEVASNLGGLPTSPQFMD-----PGLALHLAGTTRIGFDKA 552
Db 457 -FTDODRALHFEITDKTEAVARKMGVSIKAGALKDYVVPYQSTHNTGTTWGADRE 515
QY 553 TTVADNNSLVWDFANLYVAGNGTIRTFGEENPTLTSCHKAISARSIIINTLKG 606
Db 516 TSVNTYLQHWADNLEVVVGAGNFQNSGYNPTDVGALAYRCAEGILKYHSG 569

RESULT 2
B38575
glucanate 2-dehydrogenase (FAD) (EC 1.1.99.3) catalytic chain precursor (validated) - EMBL
C:Species: Erwinia cyrripedii
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 08-Dec-2000
C:Accession: B38575; T52554
R:Yun, D.Y.; Lee, Y.P.; Pan, J.G.
J. Bacteriol. 179, 6366-6372, 1997
A:Title: Cloning and expression of a gene cluster encoding three subunits of membrane-bound 2-dehydrogenase (FAD) from *Erwinia cyrripedii*
A:Reference number: A38575; MUID:98012950
A:Accession: B38575
A:Molecule type: DNA
A:Residues: 1-615 <YUN>
A:Cross-references: GB:097665; NID:g2584859; PIDN:AAC45885.1; PID:g2584862
A:Experimental source: strain ATCC 29267
A:Note: part of this sequence, including the amino end of the mature protein, was determined from GB/EMBL/DDBJ
A:Accession: T52554
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-615 <YUN>
A:Cross-references: EMBL:U97665; PIDN:AAC45885.1
A:Experimental source: ATCC29267
C:Genetics:
A:Start codon: GTG
C:Complex: heterotrimer, consisting out of cytochrome c chain (T52555), catalytic chain
C:Function:
A:Description: EC 1.1.99.3 [validated, MUID:98012950]
C:Keywords: FAD; oxidoreductase
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-615/Product: gluconate 2-dehydrogenase catalytic chain #status experimental <MAT>

Query Match 5.8%; Score 191.5; DB 2; Length 615;
Best Local Similarity 21.1%; Pred. No. 1.9e-06;
Matches 134; Conservative 83; Mismatches 256; Indels 163; Gaps 32;

QY 36 VDVFATAGSPIGATYAKLCVEAGLRVWVEIGA-----ADSEYAVNAEGTAVPYPGYH 90
Db 29 DVAVVGFAGWAGIMAKELTEAGLVNVALERGHRTYDPGAYPOSIDELT-----YN 81
QY 91 KKNTEFEQKIDIRFNVNATKGAQQVSPVRNQNVTLPDGAWSAPGSSAISGNKPNQHR 150
Db 82 IRKLL-FQ-DLSKSVTIRHDSQATVPR----- 109
QY 151 EFENLSAEAVTRGVGMSTHWTCTSPRIHPMESLPG-----IGRPKLSNDPAEDD--KE 203
Db 110 ---QLAAFLPGGTGGAGLHNSGVHFRVDPVELNLSRHYEARVGNFPEGMTIQDFGVS 166
QY 204 WNEL---YSEARLLGTSTKEFDEIRHLVLSQDAY--KDRQIRPLP-----LAC 253
Db 167 YNELEPFDDQAEKVFGTGSSAW--TIKMGKRGKGGNFYAPDRSSDF-PLPAQRRTYSA 223
QY 254 HRLKNAPEVVEWH-----SAENLFHSTYND----- 279
Db 224 QLFAQAASVGYHPYDMPSPANTSGYTNTYGAQMPCNFCGCSGYACYWYSKASPNVNI 283

QY 280 -----RQKKLFTLLNHRCTRLALTGYEKKIGA-----AEVRNLLATRNPSO--LDSYM 329
Db 284 LPALRQEPKFLRNNAVYLRVNLTKDKKRACTGYLDGGREVV--QPADLVILSAFQ 340
QY 330 AKVYVILASGAINP-OILYNSGFSGLQVTPRNDSLIPNL-GRYITEQPM--AFQCIIVLR 384
Db 341 HNVHMLLSGIGOPYNPITNEGVGRNFAYQNISTLKALFDKNTTNPFIAGGAGVAYD 400
QY 385 QEFVDSVRDDPYGL-----PWKEAVAQIAKNPTDALPIPRDPDPQVTPPTTEHPHWH 440
Db 401 DFNADNDHGPIGVGGSPFW-----VNOAGTRPVSGLPTPKGTPN-----WGS 444
QY 441 QIHRDAFSYGAVGPEVDSRVIVDLRFEAGTDPANLLVFNQVQDGYSP--QPTFRYR 498
Db 445 QWK-----AAVADTYNHHISMDAH--GAHQSYRANYLQDLPNKNYVQGLLRWTFDQ 496
QY 499 PSTASNVR-ARKMADMCVEV--ASN-----LGYLTPSPQFMDPGL--ALHLAGTTRIGF 549
Db 497 D---NDIRMAOPMVGMKMKRITAMNPKMTIGG--AKGPGTHFTTYYQTHMSGGAIMG 551
QY 550 DKATTVADNNSLVWDFANLYVAGNGTIRTFGEENPT 585
Db 552 DPKTSANRYLQSDVPNVFVGASAFPOGLGYNPT 587

RESULT 3
H87451
oxidoreductase, GMC family CC1634 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: H87451
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg
N. J.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; K
n. J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser,
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of *Caulobacter crescentus*.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: H87451
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-579 <STO>
A:Cross-references: GB:AE005673; NID:gl3423038; PIDN:AAK23612.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC1634

Query Match 5.8%; Score 190.5; DB 2; Length 579;
Best Local Similarity 20.2%; Pred. No. 2.1e-06;
Matches 136; Conservative 83; Mismatches 270; Indels 185; Gaps 28;

QY 18 INQIQPKNAIHETVNDGVDVFIAGSGPIGATYAKLCVEAGLRVWVNAEGAADSFYAVNA 77
Db 4 LNGRRRK--NTY-----DAIVGSGITGTAAKELTEKGLKVLV----- 42
QY 78 EGTAVPYVPGYHKKNEIEFQKIDIRFNVNIGALQQVSPVRNQNVT--LDPGWSAP 135
Db 43 ERGPMVRLHEDY-----PTAMLDPMQSKYP 67
QY 136 PGSSA-----ISNGKNPHORE--PENLSAEAVTRG--VG 165
Db 68 QOKLPEALNAHYKQVRRGYNTGYMTQTQFFVRDDHDEPHYTEENRFDWI-----RGYHV 121
QY 166 GMSTHWTCTSPRIHPMESLPGRPKLSND---PAEDDKENNELYSEARLIGTSTKEF 222
Db 122 GRSLLWGRQSYR-HSPID-FEANAREGIADVPIRYEDLAPW--YEHVERFIGVSGQA- 175
QY 223 DESIRH-----TLVLSQDAYKDRORI-FRPLPLACHLKNAPVEVHSA----- 268
Db 176 -EGLPHFDGHYQPPMELNCVEKAFKARSEAFRRVIRTAHLTDPTERQALGRTK 234
QY 269 ---ENL-----FHSIYNDKOKKFLTLTNHRCFLALTGG-----YEKKIGAAEVR 312

Search completed: September 26, 2002, 11:11:41
Job time: 258 sec

```

Query Match      2.9%; Score 96.5; DB 1; Length 684;
Best Local Similarity 19.2%; Pred. No. 16;
Matches 123; Conservative 61; Mismatches 224; Indels 233; Gaps 27;

QY   37 DVFIAGSGPIGATYAKLCVZBAGLRVWVEICGAADSFTAVNAEETAVPYVPGYHKKNIE 96
    : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   203 DAVVVGSGAGGAWARTLARGLDWWLEGR---WTV-----EEFR 242

QY   97 FQRDIDRFVNKALQOVSPVRNQNVPTLDPGAWSAPPGSSAISNGKNPHQREFENLS 156
    : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   243 STHPVDRYAGLYRGAGATVA-----LGRFAVLPMG----- 273

QY   157 AEAVTRGVGGMS---THTWCSTPRHPMPESLPGTGRPKLSNDPADDDKEWNELYSEAERL 214
    : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   274 -----RAVGTTVVNSGTCFPPSL-----AVQRWRDEFGGLADPDQLGRRL-DDAEQT 322

QY   215 IGTSTKFDESIRTLVLRSLDQAYDKRQRIFRPLPLACHLNKPAN----- 260
    : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   323 LRVAVPVLEIMGRNG---RLLLQAASLSUGRAAPIP-----RNAPGCGCCQCAIGCPS 373

QY   261 -----EVVEVHSANLENPHSYINDDKOKKLFTLTNNHCRTIALTG 301
    : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   374 NAKFGVHLNALPOACAAGARIISWARVERILHR----- 406

```

| | | |
|----|-----|--|
| QY | 410 | AKNFTDALP--IPRDPPEPQVTP-----FTBEHPWHT-----QIHRDAFS-YGAVGPEVD 457 |
| Db | 292 | SRLLPTAVTEPMPEQRIGGVSFTPAVFGLSLAEDWTNRGDLMAQWRLCGSYIGMIRP--- 348 |
| QY | 458 | SRVIVDLRFW-GATDPEANLLVFQNDVQGYSNPQP-TFYYRSTASNVARKWADMC 515 |
| Db | 349 | -RGVGSVRPLPGINEP-----LVSEKLAPEDWISLGVLTLGQAMFAAG--ARKVIP--- 398 |
| QY | 516 | EVASNLAGYLPSPQPMPDGL-----ALHLAGTRIGRDKATTIVADNNSLVDFAN 567 |
| Db | 399 | SISGHEGWTNDEVDERKNKLPKATNLMTHLFSCTPCGHRDACAVDSYGRVRGVEN 458 |
| QY | 568 | LYVAGNGTIRGFGENPFLTSMCHAIKARSIIIN 601 |
| Db | 459 | LFVADGVSIPAPGVNQMTMALAFRIAEALS 492 |

| | |
|------------|--|
| RESULT | 13 |
| W70T_HUMAN | |
| ID | W70T_HUMAN |
| STANDARD | PRT; 925 AA. |
| AC | P57737; |
| DT | 16-OCT-2001 (Rel. 40, Created) |
| DT | 16-OCT-2001 (Rel. 40, Last sequence update) |
| DT | 01-MAR-2002 (Rel. 41, Last annotation update) |
| DE | 70 kDa WD-repeat tumor rejection antigen homology. |
| OS | Homo sapiens (Human). |
| OC | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. |
| NCBI_TaxID | 9606; |
| RN | [1] |
| RP | SEQUENCE FROM N.A. |
| RA | Kawabata A., Hikiuji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S., |
| RA | Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T., |
| RA | Tanaka T., Nakamura Y., Isogai T., Sugano S.; |
| RL | "NEDO human cDNA sequencing project." |
| RT | Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases. |
| CC | -!- SIMILARITY: CONTAINS 8 WD REPEATS (TRP-ASP DOMAINS). |
| CC | -!- SIMILARITY: BELONGS TO THE CORONIN FAMILY OF WD-REPEAT PROTEINS. |
| CC | THIS SWISS-PROT entry is copyright. It is produced through a collaboration |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation |
| CC | the European Bioinformatics Institute. There are no restrictions on its |
| CC | use by non-profit institutions as long as its content is in no way |
| CC | modified and this statement is not removed. Usage by and for commercial |
| CC | entities requires a license agreement (see http://www.isb-sib.ch/announce/) |
| CC | or send an email to license@isb-sib.ch . |
| CC | ----- |
| CC | ENBL; AKO25674; BAB15211.1; - |
| DR | InterPro; IPR001680; WD40. |
| DR | Pfam; PF00400; WD40; 6 |
| DR | PRINTS; PR00320; GPROTEINRPT.. |
| DR | SMART; SM00320; WD40; 6. |
| DR | PROSITE; PS00678; WD_REPEATS_1; 1. |
| DR | PROSITE; PS50082; WD_REPEATS_2; 4. |
| DR | PROSITE; PS50294; WD_REPEATS_REGION; 2. |
| KW | Repeat; WD repeat. |
| FT | REPEAT 75 115 WD 1. |
| FT | REPEAT 124 163 WD 2. |
| FT | REPEAT 166 205 WD 3. |
| FT | REPEAT 209 253 WD 4. |
| FT | REPEAT 542 582 WD 5. |
| FT | REPEAT 592 632 WD 6. |
| FT | REPEAT 635 674 WD 7. |
| FT | REPEAT 728 768 WD 8. |
| SQ | SEQUENCE 925 AA; 100574 MW; A393CE973C94FA9 CRC64; |

| | |
|-----------------------|--|
| Query Match | 3.08; Score 98; DB 1; Length 925; |
| Best Local Similarity | 20.28; Pred.No. 19; |
| Matches | 88; Conservative 139; Indels 156; Gaps 21; |

| | | |
|----|-----|--|
| QY | 129 | PGAWSAPPGSSAINGNKNPHOREFENISAEAVTRGV-----GGMSTHWTCSTPIRHPPME 183 |
|----|-----|--|

RL Adv. Dent. Res. 10:111-118(1996).
CC - FUNCTION: PEPTIDES DERIVED FROM THE PARENT ENAMELIN ARE COMPONENTS
CC OF ENAMEL, A UNIQUE AND HIGHLY MINERALIZED ECTODERMAL TISSUE
CC COVERING VERTEBRATE TEETH.
CC
CC - TISSUE SPECIFICITY: EXPRESSED BY SECRETORY-PHASE AMELOBLASTS.
CC TO THE MOST SUPERFICIAL LAYER OF THE DEVELOPING ENAMEL MATRIX,
CC INTACT ENAMELIN AND LARGE-MOLECULAR-WEIGHT ENAMELINS ARE LIMITED
CC WHILE LOW-MOLECULAR-WEIGHT ENAMELINS ARE OBSERVED IN DEEPER
CC ENAMELIN. PREFERENTIAL LOCALIZATION AMONG THE CRYSTALLITES IN ROD
CC AND INTERROD ENAMEL.
CC
CC - DEVELOPMENTAL STAGE: EXPRESSED FROM LATE DIFFERENTIATION TO THE
CC TRANSITION STAGE.
CC
CC - PTM: PARENT ENAMELIN IS PROTEOLYTICALLY CLEAVED INTO SEVERAL
CC SMALLER POLYPEPTIDES. CLEAVAGE OF N-TERMINAL REGION OF ENAMELIN
CC OCCURS SOON AFTER SECRETION.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U52196; AAD10837.1; -
CC GlycosultedB; O97939; -
CC Signal; Enamel; Glycoprotein; Hydroxylation; Phosphorylation.
CC
CC SIGNAL 1 38
CC CHAIN 39 1142 ENAMELIN.
CC CHAIN 39 ? 56 KDA ENAMELIN.
CC CHAIN 39 665 89 KDA ENAMELIN.
CC CHAIN 39 ? 142 KDA ENAMELIN.
CC CHAIN 39 ? 155 KDA ENAMELIN.
CC CHAIN 174 276 32 KDA ENAMELIN.
CC CHAIN 515 665 25 KDA ENAMELIN.
CC CHAIN 670 ? 34 KDA ENAMELIN.
CC CHAIN ? ? 45 KDA ENAMELIN.
CC MOD_RES 53 53 PHOSPHORYLATION (PROBABLE).
CC MOD_RES 191 191 PHOSPHORYLATION.
CC MOD_RES 216 216 PHOSPHORYLATION.
CC MOD_RES 547 547 HYDROXYLATION.
CC CARBOHYD 245 252 N-LINKED (GLCNAC. . .).
CC CARBOHYD 252 252 N-LINKED (GLCNAC. . .).
CC CARBOHYD 264 264 N-LINKED (GLCNAC. . .).
CC CARBOHYD 291 291 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 462 462 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 929 929 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 1040 1040 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CONFLICT 680 680 H -> D (IN REF. 2).
CC CONFLICT 838 840 RDH -> TTI (IN REF. 2).
CC SEQUENCE 1142 AA; 128352 MW; 938306BC87CC5FC6 CRC64;

Query Match 3.0%; Score 100; DB 1; Length 1142;
Best Local Similarity 19.2%; Pred. No. 18;
Matches 118; Conservative 65; Mismatches 210; Indels 220; Gaps 32;

QY 76 NAEAGTAVYVP--GYH-----KKEIEFQKIDR-----FVNVIK 110
DB 192 NEEGGN--PYFGFGYRGFGRPYYISEEMFEQDFEKPKEPKTETPATESVN--- 245

QY 111 ALQOVSVPYRNQVPL-DP-GAWSAPPGSSAISNGKNPHOR-----EFENLSAEAVTR 162
DB 246 -----TTVPETNSTQPNAPNPRNGDTSPTGTS--GQGNPNSNPTGQNGPAPVNSGQGVPR 299

QY 163 -----GVGGMSTHTCSTPRI-----HPPWESLPGIGR---PKLSDNPAEDDKW 204
DB 300 SQSPGWGPROTIHENTPNINRGFPARRQWRPPG---PAMGHRNGFPYRNQIQGPRW 356

QY 205 NELYSAEERLIGTSTKEFDESIRHTLVLSQDAYKDRQIRFPLPLACHRLKNAPYVE 264
DB 357 NSFTLECK-----QAVRPGYTYRIVY-----GSTARSNPPNYA- 390

QY 265 WHSAENLFHSIYNDKOKKLTLLTNHRCRLALTG-----GYEKKIGAAE 310
DB 391 -GNSANLRRRPEGNK-----NPMVTNVAPPGKGTGVDQNMENIQNPREKQVSQKE 440

QY 311 VRNLATRNPS-----SOLDSYIMAKVYVLASGAIGNPQILYNSGFSGLQVTPRNDLSLPN 366
DB 441 -RTVVPTRDPSGWRNSQDYINKSNYKL-----PO-----PEDNMLVPN 479

QY 367 LG-----RYITEQPMACQVLVRQEFVDSVRDDPYGLPWWKEA---VAQ 407
DB 480 FNSIDQRENSYPRGESKRAPNSDQGTQIIPK-----GIVLEPRPIPYESETNQPELK 534

QY 408 HIAKNP--TDALPIPPRDEP-----QVTPPTEHPHWTQIHRDAFSGVAGVPEV 456
DB 535 HSAYPQVYITEGIPSPAKEHFPAGRTWNQOEISPPFKED-PGROEHLPLSHG----- 587

QY 457 DSRVIDLRFVFGATDPEANLLVFQNDVQDYSMP-----OPTFRYRSTASNVRRKMA 512
DB 588 -SRVHYVYDYNPDRENPSYLRNSWTYERDSDSPNTMGQENPHYPWNTDPKET----- 642

QY 513 DMCEVASNLGGYLTSPQPFMDPLALHLAGTTRIGFDKATTVADNNSLVYDFANLYVAG 572
DB 643 -----IPYNEEDPIDTGDHFPGQSR-----WDMBELSFKE 674

QY 573 NGTIRTGFGENPT 585
DB 675 DPTVRHYEGEOYT 687

RESULT 10
MTLD_KLEPN STANDARD; PRT; 382 AA.
AC O9XBNG;
DT 30-MAY-2000 (Rel. 39, Created).
DT 30-MAY-2000 (Rel. 39, Last sequence update).
DT 16-OCT-2001 (Rel. 40, Last annotation update).
DE Mannitol-1-phosphate 5-dehydrogenase (EC 1.1.1.17).
GN MTLD.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RAY 2026;
RA Otte S., Lengeler J.W.;
RT "The mtI genes and the mannitol-1-phosphate-dehydrogenase from
RT Klebsiella pneumoniae";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC - CATALYTIC ACTIVITY: D-mannitol 1-phosphate + NAD(+) -> D-fructose
CC 6-phosphate + NADH.
CC - SIMILARITY: BELONGS TO THE MANNITOL DEHYDROGENASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF166095; AAD45386.1; -
CC InterPro; IPR000669; Mannitol_dh.
CC Pfam; PF01232; Mannitol_dh; 1.
CC PRINTS; PRO0084; MTLDDHGNASE.
CC PROSITE; PS00974; MANNITOL_DHGENASE; 1.
KW Oxidoreductase; NAD.
FT NP_BIND 3 14 NAD (BY SIMILARITY).
SQ SEQUENCE 382 AA; 41130 MW; D9927E6C068C5344 CRC64;

Query Match 3.0%; Score 98.5; DB 1; Length 382;
Best Local Similarity 22.0%; Pred. No. 4.6;

RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: REQUIRED FOR BOTH FULL TRANSCRIPTION REPRESSION AND
CC ACTIVATION OF MANY GENES INCLUDING CELL TYPE-SPECIFIC GENES (STB6,
CC TY2 AND HO), CELL DIFFERENTIATION-SPECIFIC GENES (SPOL3), GENES
CC THAT RESPOND TO EXTERNAL SIGNALS (PHO5) AND TRK2. IT IS PROBABLY
CC INVOLVED IN THE SAME TRANSCRIPTIONAL REGULATORY FUNCTION OR
CC PATHWAY AS THE TRANSCRIPTIONAL REGULATORY PROTEIN RPD3.
CC -!- SUBUNIT: IT PROBABLY FORMS A COMPLEX WITH THE TRANSCRIPTIONAL
CC REGULATORY PROTEIN RPD3.
CC -!- SUBCELLULAR LOCATION: NUCLEAR, POSSIBLY LINKED TO CENTROMERE.
CC -!- DOMAIN: CONTAINS A PAIRED AMPHIPATHIC HELIX MOTIFS, SEPARATED BY A
CC 10 TO 30 AA SEGMENT THAT FORMS POSSIBLY A LOOP, RESULTING IN A
CC STRUCTURE SIMILAR TO THAT OF HLH AND TPR MOTIFS.
CC -!- SIMILARITY: TO S.POMBE SPAC12C2.10C.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M36822; AAA34839.1; -
CC EMBL; 274746; CA999003.1; -
CC PIR; S12068; RGYB33.
CC SGD; S0005364; SIN3.
CC InterPro; IPR003822; PAH.
CC Pfam; PF02671; PAH; 3.
CC KW Repeat; Transcription regulation; Repressor; Cell division;
CC Activator; Nuclear protein.
CC FT DOMAIN 239 1200 4 X 2 PAIRED (A,B) AMPHIPATHIC HELICES.
FT REPEAT 239 252
FT REPEAT 274 285
FT REPEAT 426 439
FT REPEAT 461 472
FT REPEAT 679 692
FT REPEAT 714 725
FT REPEAT 1152 1165
FT REPEAT 1189 1200
FT REPEAT 480 519
FT CONFLICT 510 510 O -> QAO (IN REF. 2).
FT SEQUENCE 1536 AA; 174838 MW; 0834726312B13878 CRC64;

Query Match 3.1%; Score 102.5; DB 1; Length 1536;
Best Local Similarity 19.3%; Pred. No. 19;
Matches 115; Conservative 83; Mismatches 192; Indels 207; Gaps 31;
QY 72 FYAVNAEETAVPVPGYHKKNEIEFQKIDRFVNV---IKGALQVSVPRVNVPTLD 128
DB 985 FYDILCLADFTTHTPAYSNPDKERLKLKYPISLFFSISFEKIEESLYSHKQNV--- 1040
QY 129 PGAWSAAPGGS---SAISNGKNPHOREFENLSA-----EAVTRGVGMSTHWTCTPRIH 179
DB 1041 ---SESSGSDGSSATSKRPYQEQMSLLDLHRSRYQKLKR-----SNDEGKVPQLS 1091
QY 180 PMWESLPG-IGRPKLSNDAEDD-----KAW 204
DB 1092 EPPEEPNTEIEBELIDEAKNPWLTGNLVEANSQIIQNRSIFNLANTNIYIFFRHW 1151
QY * 205 NELYSAEALIGTSTEFDESIRHTLVLSL-----QAYKD 241
DB 1152 TTIY-----ERLL--EIKOMNERTVKEINTRSTVTFAKDLILLSQLSEMGLDFYGEDAYKQ 1206
QY 242 RQIRFPLPLACHRLNAPEYVWEHSAENLFHSIYNDKOKKLTFL-----LTNHRCTR 295
DB 1207. VLRLSR-----RLINGDLEHQWF--EESLROAYN-NKAFKIYTTDKVTQSLVKAHT- 1255
QY 296 LALTGYEKKIGAAEVRNLLATNPSSQLDSYIMAKVYVLASGAIGNPQILYNSGFGSLG 355
DB 1256 -LWTDAKTAEINALFYKD-----RN-----ASTSAKDQIIYR-----LQ 1289

QY 356 VTPRNDSLIPNLGR-----YITEQPMAFQOIVLRQEFVDSVRDDPYGLPMWKEAVAQ 407
DB 1290 VR-SHMSNTENMPRIEDFKRTLHVSIOYIALDLTLKEPKADESK-----WKYVVT 1340
QY 408 HIAKNTDIALPIFRDEPOVTPFTTEHHPWHTQIHRDAFSGYAGVDSRVIVDLRWF 467
DB 1341 YALPHPTTEGI-----PHEKLKIPFLE-----RLIEFGQDIDGTEVDEEF----- 1379
QY 468 GATDPEANNL-----LVFQNDVQDG-YSM--POPTFRYPSTASNVRAKMMADM 515
DB 1380 ---SPEGISVSTLKIKIQTPIYQIHLIENGSDYVTRKATNKY-PTIANDNTQKGMYSOKK 1435
QY 516 EVASNLGGYLPTSPQFMD--PGLALHLACTTRIGEDK-----ATTVAONNSL 561
DB 1436 ELIS-----KFLDCAVGLRNRLDEAOKLSMOKWENLKDSIAKTSAGNOGI 1481
RESULT 7
SCA4_RICPA STANDARD; PRT; 1010 AA.
AC O9AJ75;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antigenic heat-stable 120 kDa protein (PS120) (120 kDa antigen)
DE (protein PS 120) (fragment).
GN SCA4 OR D.
OS Rickettsia parkeri.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_Taxid:35792;
RN [1]
RP SEQUENCE FROM N.A.
RA Sekeyova Z., Roux V., Raoult D.;
RT "Phylogenetic analysis of Rickettsia spp. by comparing sequence of the
RT 'gene D' coding for an intracytoplasmic protein."
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF155059; AAK30690.1; -
KW Antigen.
FT NON_TER 1010 1010
FT SEQUENCE 1010 AA; 110694 MW; 78586D8C92FF9C5B CRC64;

Query Match 3.1%; Score 101; DB 1; Length 1010;
Best Local Similarity 18.2%; Pred. No. 13;
Matches 106; Conservative 88; Mismatches 243; Indels 144; Gaps 25;
QY 136 PGSAISNGKNPHOREFENLSAEAVTRGVGMSTHWTCTSTPR---IHPPMESLPGIGRP 191
DB 55 PMSVLSGNISPSQ-----TSDPTKAV-----RETIQPKDNLIEQLKDLAAL--- 100
QY 192 KLSNDPAEDDKENLYSEAEERLIGT-----STKEF-DESIRHTLVLSLQD-----AYKD 241
DB 101 -TORDLAEOQRKEIEBEKEKDKTLSTFFGPNANREFIDKALENPELKKELESIAGYN 159
QY 242 RQIRFPLPLACHRLNAPEYVWEH---SAENLFHSIYNDKOKKLTFL----- 207
DB 160 VHNFTS-----AASGYPGGKPVQVQWENHVSASDLRATVYNDAGDELCTNETTVTKPFT 215
QY 288 LTNHRTALALTGYE-----KKIGAAEVRNLLATR-----NPSSQLDSYIMAKVYVL 335
DB 216 LAKQDGTQVOISSYREIDFPFKDKADGSMHL-SWVALKADGTPKSKRPVFTAH---Y 271

OS Aspergillus oryzae.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 OX NCBI_TaxID=5062;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91254744; PubMed=1368680;
 RA Hata Y., Kitamoto K., Gomi K., Kumagai C., Tamura G., Hara S.;
 RT "The glucoamylase cDNA from Aspergillus oryzae: its cloning,
 RT nucleotide sequence, and expression in Saccharomyces cerevisiae.";
 RL Agric. Biol. Chem. 55:941-949(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92104497; PubMed=1761224;
 RA Hata Y., Tsuchiya K., Kitamoto K., Gomi K., Kumagai C., Tamura G.,
 RA Hara S.;
 RT "Nucleotide sequence and expression of the glucoamylase-encoding gene
 RT (gluA) from Aspergillus oryzae.";
 RL Gene 108:145-150(1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIB 40;
 RA Hara S., Tamura G., Kumagai C., Gomi K., Kitamoto K., Tsuchiya K.,
 RA Hata Y.;
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
 CC glucose residues successively from non-reducing ends of the chains
 CC with release of beta-D-glucose.
 CC -1- SIMILARITY: BELONGS TO FAMILY 15 OF GLYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL; D01035; BAA00841.1; -
 DR EMBL; D10698; BAA01540.1; -
 DR PIR; J01346; J01346.
 DR HSP; P04064; IGAI.
 DR InterPro; IPR002044; CBD_4.
 DR InterPro; IPR000165; Glyco_hydro_15.
 DR Pfam; PF00686; CBD_4; 1.
 DR Pfam; PF00723; Glyco_hydro_15; 1.
 DR PRINTS; PR007136; GLHYDRLASE15.
 DR ProDom; PD001368; CBD_4; 1.
 DR PROSITE; PS00820; GLUCOAMYLASE; 1.
 KW Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein;
 KW Signal.
 FT SIGNAL 1 19 BY SIMILARITY.
 FT PROPEP 20 25 BY SIMILARITY.
 FT CHAIN 26 612 GLUCOAMYLASE.
 FT BINDING 146 146 SUBSTRATE (BY SIMILARITY).
 FT ACT_SITE 202 202 CATALYTIC BASE (BY SIMILARITY).
 FT ACT_SITE 205 205 GENERAL ACID CATALYST (BY SIMILARITY).
 FT ACT_SITE 206 206 INTERACT WITH SUBSTRATES (BY SIMILARITY).
 FT DISULFID 236 239 BY SIMILARITY.
 FT DISULFID 248 475 BY SIMILARITY.
 FT DISULFID 288 296 BY SIMILARITY.
 FT CARBOHYD 39 39 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 612 AA; 65486 MW; CD7B23E5FA978F97 CRG64;
 Query Match 3.2%; Score 103.5; DB 1; Length 612;
 Best Local Similarity 19.0%; Pred. No. 3.9;
 Matches 125; Conservative 67; Mismatches 207; Indels 259; Gaps 31;
 QY 3 LSTEQMLRDYPRSMQINGQIPKNAIHETVGNCDGVDFIAGSGPIGATYAKLCVEAGLRVV 62
 DB 34 LSTE-----ANFSRAILNITGADGQSAQSGAPGVVIASPSK----- 70

QY 63 MVEIGADSFYAVNAEGTAVPVPGYHKKNEIEFQKIDRFVNVIKGALQOVSPVRNQ 122
 DB 71 ----SDPDYFTWRDSDLVGLVMKTLVDFRGGDADLLPIIEEFIS-SQARIQGISNP---- 121
 QY 123 NVPTLDPGAWSAPPGSSAISNGKNPHQREFENLSAAVTRGVGGMSTHWTCTPRTPHPM 182
 DB 122 -----SGALSSG----- 128
 QY 183 ESLPGIGRKLSDNPAEDDKENWELYSERLIGTSTKEFDESI----RHTVLRSIQDAY 239
 DB 129 ----GLGEKFNVDFTAGMORPORDGAPALRATANISFGEHLVENSHTSIATDL----- 180
 QY 240 KDRQRIERPLPLACHRLKNAPEYVEHMSAENLFHSIYNDKQKLFLETLNHRCTRALT 299
 DB 181 -----VMPV-----VRNDSLYAQWQSGF-DLWEVQGTSEFTVAVSHR-----ALV 223
 QY 300 GG--YEKKTG-----AAEVRNLLATRNPSQLDSYIMAKVYVLSAIGNPOILYN 348
 DB 224 EGSSFAKTVGSSCPYCDQAPQVRCYL---QSFMTGSIQAN-----EG 264
 QY 349 SGFSGIQVTPRNDLIPNLGRYTEOPMAFCQIVLRQEFVDSVRDDPYGLPMWKEAVAOH 408
 DB 265 GGRSGKDIN-----TVLGSHTFDQATC-----DDATFQPCSRALANH 304
 QY 409 IAKNPTDALPIPE-----RDPEQVTTPTTEHPWH-----TOIHRDA 446
 DB 305 --KVVTDSERSIIVAINSGRAENQAVAVGRYPEDS---YVNGNPWELTTLAAAEQLYDAL 358
 QY 447 FSVGAVGPEVDSRVYD--LRWF-----GATDPEANLLYFONDVQ-----DGYSMPQ 492
 DB 359 YQMDKGTG---SLAIVDVLSPFKALYSAAATGYASTTVYKDIYSAVKAYADGVQIV 414
 QY 493 PTFPRYPTSTASNVARKMADMCVASNLGGYLPSPQFMDPGLALHLAGTTRIGFDKA 552
 DB 415 QT--YAASGTG-----MAEQ-----YKTDGSO----- 435
 QY 553 TTVADNNSLVMDPANLYVAG---NGTIRTFGEN-----PTLSMCHAIKSARSINT 602
 DB 436 TSARD---LTWSYAALITANNRRNAVVPAPWGETAATSIPTASCTTSASGTYSVVIT 490
 RESULT 5
 AMYG_ASAPK STANDARD; PRT; 639 AA.
 AC P23176;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Glucoamylase I precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)
 DE (1,4-alpha-D-glucan glucohydrolase).
 GN GAI.
 OS Aspergillus awamori (var. kawachi).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 OX NCBI_TaxID=105351;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Hayashida S., Kuroda K., Ohta K., Kuhara S., Fukuda K., Sakaki Y.;
 RT "Molecular cloning of the glucoamylase I gene of Aspergillus awamori
 RT var. kawachi for localization of the raw-starch-affinity site.";
 RL Agric. Biol. Chem. 53:135-141(1989).
 RN [2]
 RP SEQUENCE OF 494-538.
 RA Hayashida S., Nakahara K., Kuroda K., Miyata T., Iwanaga S.;
 RT "Structure of the raw-starch-affinity site on the Aspergillus awamori
 RT var. kawachi glucoamylase I molecule.";
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
 CC glucose residues successively from non-reducing ends of the chains
 CC with release of beta-D-glucose.
 CC -1- SIMILARITY: BELONGS TO FAMILY 15 OF GLYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration

Db 727 VKYFGTNLFTVDAGIIPHLPTG---NPQGTLMSAAEQAAAKIL-ALAGG 772

RESULT 2
G6PD_MYCTU

| ID | G6PD_MYCTU | STANDARD; | PRT; | 514 AA. |
|----------|--|-----------|------|---------|
| AC | O08407; | | | |
| DT | 30-MAY-2000 (Rel. 39, Created) | | | |
| DT | 30-MAY-2000 (Rel. 39, Last sequence update) | | | |
| DT | 16-OCT-2001 (Rel. 40, Last annotation update) | | | |
| DE | Glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49) (G6PD). | | | |
| GN | ZWF OR ZWF2 OR RV1447C OR MT1494 OR MTCY493.07. | | | |
| OS | Mycobacterium tuberculosis. | | | |
| OC | Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; | | | |
| OC | Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium. | | | |
| ON | NCBI_TaxID=1773; | | | |
| OX | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=H37RV; | | | |
| RX | MEDLINE=98295987; PubMed=9634230; | | | |
| RA | Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., | | | |
| RA | Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F., | | | |
| RA | Badcock K., Basham D., Brown D., Chillingworth T., Connor R., | | | |
| RA | Davies R., Davlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S., | | | |
| RA | Hornsby T., Javels K., Krogh A., McLean J., Moule S., Murphy L., | | | |
| RA | Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., | | | |
| RA | Rutter S., Seeger K., Skelton S., Squares S., Squares R., | | | |
| RA | Sulston J.E., Taylor K., Whitehead S., Barrell B.G.; | | | |
| RT | "Deciphering the biology of Mycobacterium tuberculosis from the | | | |
| RT | complete genome sequence." | | | |
| RL | Nature 393:537-544(1998). | | | |
| [2] | | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=CDC 1551 / Oshkosh; | | | |
| RA | Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O., | | | |
| RA | Peterson J.F., Deboy R., Dodson R., Ginn M.L., Haft D., Hickey E., | | | |
| RA | Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L., | | | |
| RA | Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A., | | | |
| RA | Bisai W.; | | | |
| RT | "Whole genome comparison of Mycobacterium tuberculosis clinical and | | | |
| RT | laboratory strains." | | | |
| EL | Submitted (APR-2001) to the EMBL/GenBank/DBJ databases. | | | |
| CC | -1- CATALYTIC ACTIVITY: D-glucose 6-phosphate + NADP(+) -> D-glucono- | | | |
| CC | 1,5-lactone 6-phosphate + NADPH. | | | |
| CC | -1- PATHWAY: FIRST STEP IN PENTOSE PHOSPHATE PATHWAY. | | | |
| CC | -1- SIMILARITY: BELONGS TO THE GLUCOSE-6-PHOSPHATE DEHYDROGENASE | | | |
| CC | FAMILY. | | | |
| CC | -1- CAUTION: M.TUBERCULOSIS HAS TWO GENES FOR ZWF. THIS ONE LOOKS LIKE | | | |
| CC | A CLASSICAL ZWF. | | | |
| CC | ----- | | | |
| CC | THIS SWISS-PROT entry is copyright. It is produced through a collaboration | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announcement/ | | | |
| CC | or send an email to license@isb-sib.ch). | | | |
| CC | ----- | | | |
| DR | EMBL; Z95844; CAB09259.1; .. | | | |
| DR | EMBL; AE007019; AAK45757.1; .. | | | |
| DR | HSSP; P11411; 1DGP. | | | |
| DR | TIGR; MT1494; .. | | | |
| DR | TubercuList; Rv1447c; .. | | | |
| DR | InterPro: IPR001282; G6PD. | | | |
| DR | Pfam; PF00479; G6PD; 1. | | | |
| DR | Pfam; PF02781; G6PD.C; 1. | | | |
| DR | PRINTS; PR00079; G6PDHGRGNASE. | | | |
| DR | PRODOM; PD001129; G6PD; 1. | | | |
| DR | PROSITE; PS00069; G6P_DEHYDROGENASE; 1. | | | |
| KW | Oxidoreductase; NADP; Glucose metabolism; Complete proteome. | | | |
| FT | ACT_SITE 210 210 BY SIMILARITY. | | | |
| SEQUENCE | 514 AA; 57343 MW; 22A9CCDC9AB062F CRC64; | | | |

Db 429 VVDANCKVHGMDNLYVGGSSVFCTSGQANPTTT 461

Search completed: September 26, 2002, 11:11:20
Job time: 262 sec

```

RT Mesorhizobium loti. ";
RL DNA Res. 7:331-338 (2000).
DR EXBL; AP003009; BAB52905.1; -.
DR InterPro; IPR000205; NAD_binding.
KW Complete proteome.
SQ SEQUENCE 499 AA; 54537 MW; 633924AED96B8F51 CRC64;

Query Match 4.7%; Score 155.5; DB 16; Length 499;
Best Local Similarity 20.7%; Pred. No. 0.00053;
Matches 129; Conservative 65; Mismatches 235; Indels 195; Gaps

Qy 37 DVFTAGSGPTGAYAKLCVEAGLRVWVEIGADSPAYVNAEGTAVPYVPGYHKKNEIE 96
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7 DIVIIGSGIGSLAYSADTGRRIIVLE-----RGEHLRDTPE 45

Qy 97 FOKIDIRFVNVIKALQOQSVVPRNQNVPTLDPGASAPPGSSAI-----SNCK--- 145
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 46 ARDDIAIFQNGFYRSSEE-----WLATDGESFLPCNYYVGGNSKFFPG 88

Qy 146 -----NPHQREFENLSAEAVTRGVGGMSTHTWCTSTPRIHP-----PMESLPGIGRP 191
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 89 AVMYRYRQEDFNPRDH-----NGGRSPGWPISYAELEPWYERAEILFGV-RG 134

Qy 192 KLSNDPAEDDKENWELYSERAEILGSTKEEDSIRHTLVLSQDAYKDRQIRFPDPL 251
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 135 DARODTPEPR--NRPY-----RYLPVPDEPAIATVRQLLQAGIHPA-----SLPL 179

Qy 252 A-----CHRLKNAPEYVEHSAENLFHSIYNDOKKFLTLTNHRCRTLALTGGYEKK 305
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 180 AIDTADWLRRAKTG-----WDAFPT--GAGKIDAEVGLPTKALEHPNATLTGANVQL 232

Qy 306 IGAAEVRNLLATRNPSQLDSYIMAKVYVLASGAINPQILYNSGSGQLQVTPRNDSLIP 365
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 233 VTDASGRVMAAVFIKDGVELSIGADVFVAGAGVSAOALLRS-----STSVYP 282

Qy 366 N-----LGR-YITEQPMAFQI-----VLROE---FVDSVRDDPYG-LPMWKEAV 405
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 283 NGLGNSSDOLGRNFMNHTTAMLAIDPFRNRTAVYOKTLGFENDFYKNDPLGSFPLGNVOL 342

Qy 406 AQHTAKNPDA-----LPIPRDPPEQVTTPTTEEHPMHT-QIHRDAFYSXGAVG---PEVD 457
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 343 LGHTIGNILKANAPLLP-----RWLAGLVARNICYGWLFISETDLPNPE 384

Qy 458 SRVIVDLRFEGATDPEANLLVFQNDYQDSYMPQPTFRYPSTASNVRRKMM--ADMC 515
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 385 SRVTI-----RNGRIYVNWVRNMGAAHETLIR-----RTRAVMREAGFP 423

Qy 516 EVASNLGGYLPSTSPQMPDPLALHLAGTTRIGFDKATTVADNNLSLVDFANLIYVAGNCT 575
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 424 VLFRTFGRKTT-----HOCGRFARLGSDPNTSVSPDCRASHDIANLIYVDASV 472

```

```

Db      424  VVLRTTGRKTTT-----HQCFTARLGSDPNTSVVSPDCRSHDIANLYVTDSV 472
QY      576  IRTGFGENPILTSMCHAIKSARSI 599
Db      473  LPTSAAVNPALTIAALKAGAARI 496

RESULT 14
ID       074253      PRELIMINARY;      PRT;      769 AA.
AC       074253;
DT       01-NOV-1998 (TREMBLrel. 08, Created)
DT       01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT       01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE       CELLOBIOSE DEHYDROGENASE (EC 1.1.99.18).
GN       CDH.
OS       Pycnopusorus cinnabarinus.
OC       Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OX       Stereales; Schizophyllaceae; Pycnopusorus.
NCBI_TaxID=5643;
[1]
RN
RP      SEQUENCE FROM N.A.
RX      MEDLINE=99321795; Pubmed=10393235;

```


QY 556 ADNNSLVWDFANLYVAGNGTIRGTGFGNPTLTSMCHAIKARSINTLK 604
 Db 502 VDGNCRTFDEHNLWLPGGGAIPSAVSNNSTLSMAALGLKRAHDISLRMK 550

RESULT 9
 Q9HOR8 PRELIMINARY; PRT; 529 AA.
 AC Q9HOR8
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE VNG1035C.
 GN VNG1035C.
 OS Halobacterium sp. (strain NRC-1).
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
 OC Halobacterium.
 OX NCBI_TaxID=64091;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20504483; PubMed=11016950;
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
 RA Shukla H.D., Lasky S.R., Balliga N.S., Thorsson V., Sbrogna J.,
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
 RA Leitthausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
 RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
 RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
 RT "Genome sequence of Halobacterium species NRC-1";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
 DR EMBL; AB005037; AAG19445.1;
 DR InterPro; IPR000425; MIP.
 DR InterPro; IPR000205; NAD_binding.
 DR PROSITE; PS00221; MIP; UNKNOWN_1.
 DR Complete proteome.
 KW SEQUENCE 529 AA; 56633 MW; FAED92051336225D CRC64;

Query Match 5.5%; Score 179.5; DB 17; Length 529;
 Best Local Similarity 22.1%; Pred. No. 6.6e-06;
 Matches 143; Conservative 53; Mismatches 236; Indels 215; Gaps 28;

QY 37 DVFIAGSGPGTATYKLCVBAGLRVWVEIGAADSFYVNAEAGTAVPYVGYHKKNEIE 96
 Db 10 DVCVIGAGPAGGLIADRLASDGHSHVVLEAG--PRFDAADR----- 50
 QY 97 FQDIDRFVNVIGALQOVSVYVNRQNVPLDPCAWS-APPGSSAISNGKNPHOREFENL 155
 Db 51 -----RMERSIRPAHGPAV-----WEMGGPRDAYASTGD-----RYPL 85
 QY 156 SAEAVTRGVGMSTHTCTSPTRTHPP---MESLPGIGRKLNDNDPAEDDKWENLYSEAE 212
 Db 86 NAARV-KGVGGSLHWQGYWYRHEQDFRASATGVGA-----DWPFYDTLKPYYAAAE 139
 QY 213 RLI---GTSKPEDESIRHTLVLSLDQAYKDRIFRPLPLACHRLKNAPEYVHSAE 269
 Db 140 SALGVSGASNDPAPPREQHPQAPPPSYSD--SLFAD--ACESLGAT----- 185
 QY 270 NLFHSYNDKQKLLFTLNHRCRLALTG-----GYE-----K 304
 Db 186 ---HSPVN-----ARLSAGRETRACVGYGTCQVPCPGAKYDATVHVDRATDAGAR 234
 QY 305 KIGAAEVRNL-----LATRNPSSOLDSTYMAKVVYLVASGAINQPOL----- 346
 Db 235 VIDEAPVORLEHDAAGDRVTGAVATPDGTHRQS-----ATEFLAAGGIEPTRLLLSDS 291
 QY 347 --YNSGF---SGLVTPRNDSLIPNLGRYTEQPMFACQIVLROEFYDSVRDDPYGLPW 401
 Db 292 DRYPDGLANSSGL-----VGRY-----FMD----- 311
 QY 402 KEVAQHIANKPTDALPIPRDPEDPQVTTFTTEHPHWHQIHRDAFSYGAV-----G 453

Db 312 -----HLFAGAGGTLDEPTFRQNHVGENT--TESHQY--DRPDGSRGAIKLEFLNYAG 360
 QY 454 PEVDSRVIVDLRWEGA-----TDPEANNLLVFQNDVQDSYSNPQP 493
 Db 361 PSAPMALSGDDWDGDMCDRIIRDASGTHIAGVGLVEOQPRPENVRVLRHPERTDVHGNPVP 420
 QY 494 TFRYRPSFASNVARKMMADMCEVASNLGGYLP-TSPPOFMDPGLALHLAGTTRIGFDKA 552
 Db 421 DVVWSLSAYERTIERANEIQREILTELGDADIEWTQVE--DTGPAFHHMGTTRMGTDPA 478
 QY 553 TTVDNNSLVWDFANLYVAGNGTIRGTGFGNPTLTSMCHAIKARSIS 599
 Db 479 ESVYDPRLRTHDLSNLSSVASSVPEPTAGAMNPTLTIAALAKAADHI 525

RESULT 10
 Q9RZ26 PRELIMINARY; PRT; 722 AA.
 AC Q9RZ26;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE GMC OXIDOREDUCTASE.
 GN DRA0127.
 OS Deinococcus radiodurans.
 OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
 OX NCBI_TaxID=1299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RI;
 RX MEDLINE=20036896; PubMed=10567266;
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 RA Vamathevan J.J., Lam P., McDonald L., Otterback T., Zaleski C.,
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;
 RT "Genome sequence of the radioresistant bacterium Deinococcus radiodurans RI";
 RL Science 286:1571-1577(1999).
 DR EMBL; AE001862; AAF12230.1;
 DR TIGR; DRA0127;
 DR InterPro; IPR001064; Crystallin.
 DR InterPro; IPR001199; Cyt.B5.
 DR InterPro; IPR000172; GMC_oxred.
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR Pfam; PF00732; GMC_oxred; 1.
 DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
 DR PROSITE; PS00191; CYTOCHROME_B5_1; UNKNOWN_1.
 DR PROSITE; PS00624; GMC_OXRED_2; 1.
 DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
 DR Complete proteome.
 KW SEQUENCE 722 AA; 75376 MW; 450DF1CD1B7596F9 CRC64;

Query Match 5.4%; Score 176; DB 16; Length 722;
 Best Local Similarity 20.9%; Pred. No. 2.1e-05;
 Matches 139; Conservative 77; Mismatches 239; Indels 210; Gaps 33;

QY 14 RSMOINGQIPKNAIHTHYGNDG-----VDVFIAGSGPGTATYKLCVBAGLRVWVEIGA- 68
 Db 186 RNSAITPYVQ-----DGEVLEADAVVGGSGGVIAARLAQAGKRVVLEAGGH 236
 QY 69 --ADSFYVNAEAGTAVPYVGYHKKNEIEFQKIDIRFNVVIGALQOVSVYVNRQNVPT 126
 Db 237 YHEAHFOGRELAAYQTLYYRGGYH-----PTADGNV-T 268
 QY 127 LDPGAAAPPSSAISNGKNPHOREFENLSAEAVTRGVGMSTHTCTPRIHPMESLP 186
 Db 269 LVAGA-----NLG-----GGSTVWNSV-----PPRDDI- 293


```
Db 154 GOATRVVGMSTHWTGATPRDREQ-----RPLLVKDDQDADDAEWDRLYTKAESYF 206
QY 216 GTSTKEPDESIRHTLVLSLQDAYKDRIFRPLPLACHRLKNAPEVVEHSAENLF--H 273
Db 207 KTGTDQPKESIRHNLVNLKLAEEYKG--QRDQOIPLAATR--RSPTFVWSSANTVDLQ 263
QY 274 SIYNDKQKKLFTLLTNHRTLRALTGGYEKKIGAAEVNRNLLATRNPSQDSY----- 327
Db 264 NRPNTDAPNERENLFPVACERV-----VRN-----TGNSETESLRIHDLI 304
QY 328 -----IMAKVYVLASGAIGNPQILYNSGSLG--QVTPRN--DSLIPNLGRVITTEQPAFC 379
Db 305 SGRFEKADVFLTAGVHNAQLVNSGFGQLGRDPANPPQLPLSLGSIITTEQSLVFC 364
QY 380 QIVLRQEFVDSVRDDPY--GLP-----WKKEVAOAHIAKNPTDAL 417
Db 365 QTVNSTELIDSVKSDMIIRGNPGDLGYSVTTPCAETNKHPDMNEKVKHMMHQEDPL 424
QY 418 PIPRDPPEQVPTTTEHHPWHTQIHRDAPSYGAVGEVDSRVIVDLRWFGATPEANNL 477
Db 425 PIPFDPPEQVPTTTEHHPWHTQIHRDAPSYGAVGEVDSRVIVDLRWFGATPEANNL 477
QY 478 LVFQNDYQDGYSPQPTFRYR--PSTASNRARKMMADMCEVASNLGGVLPSTPPQFMDPG 536
Db 485 LWFSDKITDYNPQPTDFRFPAGRTSKAEDMTDMCVMSAKIGGFLPGSLFQFMEPG 544
QY 537 LAHLACTTRIGEDKA--TTVADNNSLVDFANLYVAGNGTIRTFGGENPILTSMCHAIA 594
Db 545 LVHLAGLTHRMGEDEQDKCCVNTDSRVFGKFLNGLCGGCGNPTAYGANPTLTAMSLAIK 604
QY 595 SANSIIN 601
Db 605 SCEYIKN 611

RESULT 2
Q9RH54 PRELIMINARY; PRT; 551 AA.
AC Q9RH54;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE DEHYDROGENASE SUBUNIT 1.
OS Pantoea agglomerans.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pantoea.
OX NCBI_TaxID=549;
RN [1]
RP SEQUENCE FROM N.A.
RA Chang K.-I., Yun D.-Y., Pan J.-G., Shin Y.-C.;
RT "Cloning and expression of a gene cluster encoding three subunits of
RT membrane-bound 2-keto-D-glucanate dehydrogenase from Erwinia herbicola
RT ATCC08111 in Escherichia coli."
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF068066; AAF21261.1;
DR InterPro; IPR000205; NAD_Binding.
SQ SEQUENCE 551 AA; 60047 MW; FE2ECCD0933DAFEB CRC64;

Query Match 6.2%; Score 203; DB 2; Length 551;
Best Local Similarity 19.9%; Pred. No. 8.7e-08;
Matches 132; Conservative 84; Mismatches 250; Indels 198; Gaps 26;

QY 24 KNAIHTETGYNDGVDFITAGSGPTGATYAKLCVEAGLRVWVMEIGAADSFYAVNAEETAV 83
Db 2 KKPVFTAGQDASADIVTSGVGGMANELVSGYSLVLEAGL----- 46
QY 84 PYVPGYHKKNEIFQKIDIDRFVNYKGAQQSVYVPRNQ---NVPTLDPGAWSAF----- 135
Db 47 -----RIDRAQAVENRN-----MPFANRAGSDFGQLYPOSKFAPAPLYF 86
QY 136 PGSSAISNGKNHOREFENLSAEATRVGCGMSTHTWCSTPRTHPP-----MESLPGIGRPK 192
Db 136 PGSSAISNGKNHOREFENLSAEATRVGCGMSTHTWCSTPRTHPP-----MESLPGIGRPK 192
```

```
Db 87 PRNNYV--NVTGNADSFQ-----QGYLRTVGCTTWHAAASCHRRHPSDFVMQSYGVGR-- 139
QY 193 LNSDPAEDDEKEM-----NEL-----YSEAERLIGTSKEDESIHRTLVLSRLQDAKDRQRI 245
Db 140 -----DWPYGYDELEPWYCKAENIGVAGP--NDPARQSPTERS----- 176
QY 246 FRPLPLACHRLKNAPEVVEHSAENLFHSIYNDKQKKL-----FTLLTNH 291
Db 177 -QPYPM-----DMVPPAHGDNYPASVYVPHGYNLPIPOGRSTRPWGERTCCGNN 226
QY 292 RCTRLATGGYEKKIGAAEVNRNLLATRNPSQDSYIM----- 329
Db 227 NCQPICTPAGMYNGIHVE-----RAERNGAVYLAEVVYKRYMDTDSNNRITAVHWDLTSGA 282
QY 330 -----AKVYVLASGAIGNPQILYNSGSLGQVTPRNDSLIPN-----LGRYTEQPMFACQ 380
Db 283 SHKATAKAFALACNGIETPRLLL-----MAANDANPANGIANASDMVGRNMDHSGFPCS 336
QY 381 IVLROEFVDSVRDDPYGLPMWKEAVAQIAKNPT--DALPIFRDPPEQVPTTTEHHPWH 439
Db 337 FLTKR-----PVN-----LGKGPAAQSSCMVGYRQD-----PRRYSAN 370
QY 440 TQI-----HRDAFSYGAAGVPEVDSRV-----IVDLRWFGATDPEANNLVFPQND 483
Db 371 KVIILNISRVTATQQAAMKGLVKGALDEIRYRAVHSVDSLSISLEPLDPDENRLTSLKT 430
QY 484 VQDGYSPQPTFRYRSTASNRARKMMADMCEVASNLGGVLPSTPPQFMDPGGLAL--HL 541
Db 431 RKDPHGLPCPDYDYDGYVRKGAESAHAQLEHI-----GOLFDAKEFTISQGLNANNHI 485
QY 542 AGTTRIGEDKATTVADNNSLVDFANLYVAGNGTIRTFGGENPILTSMCHAIAKARSIIIN 601
Db 486 MGVINGKNAKEAYVDCNCRADFHENLWLPGGGAIPSAVYVNSTLTMAALGLKAHDISL 545
QY 602 TLKG 605
Db 546 RMKG 549

RESULT 3
Q9PI90 PRELIMINARY; PRT; 573 AA.
AC Q9PI90;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PUTATIVE OXIDOREDUCTASE SUBUNIT.
GN C30415.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences."
RL Nature 403:665-668(2000).
DR EMBL; AL139075; CAB74251.1;
KW Complete proteome.
SQ SEQUENCE 573 AA; 63686 MW; 7BE8522A779A5F94 CRC64;

Query Match 6.1%; Score 200.5; DB 16; Length 573;
Best Local Similarity 20.8%; Pred. No. 1.5e-07;
Matches 135; Conservative 74; Mismatches 270; Indels 175; Gaps 26;
```